MICROBIAL ECOLOGY OF THE IBERIAN PYRITE BELT DEEP SUBSURFACE

by

Fernando Puente Sánchez

Thesis

submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Science
Ph.D. Programme in Microbiology

Madrid, 2016

Abstract

The existence, characteristics and potential limits of the deep subsurface biosphere have been a matter of growing interest over the last decades as it poses unique challenges for the development and maintenance of life. Albeit preliminary studies made clear that the microbial abundance and diversity in the deep subsurface biosphere were much higher than originally anticipated, it remains almost completely unexplored due to the difficulty of obtaining reliable samples. Therefore, a series of wide-ranging research questions regarding the energetics, phylogenetic, metabolic diversity and ultimate origin of such ecosystems are still unanswered.

The Iberian Pyrite Belt (IPB, southwest of Spain) is the most important volcanogenic massive sulphide district in the world and the largest sulphur anomaly in the Earth's crust. It has been long hypothesized that its underground ore bodies are responsible for the unique conditions of the extremely acidic Tinto River, which is born near the sulphide-rich Peña de Hierro mining site. Moreover, the IPB deep subsurface has been recognized as a terrestrial Mars analogue due to its geological similarities to known Martian environments. Its study can thus help to elucidate several key questions, including the degree of influence of deep subsurface life in surfacial events, the possibility of a lithoautotrophic early Earth, or the potential of life as we know it for thriving in apparently barren environments such as the present Mars.

During the course of this thesis we have focused in the microbiological characterization of two boreholes (160 and 613 metres deep) drilled in the Peña de Hierro area, using both conventional and molecular ecology techniques. This has included the development of an optimized protocol for sterile DNA isolation and amplification from mineral samples, as well as the use of novel bioinformatic tools to minimize errors in 16S rRNA gene-based population studies and assess the degree of contamination introduced during sample retrieval and processing. The coupling of microbial taxonomy profiles with physicochemical data allowed us to describe the geomicrobiological processes occurring in the deep subsurface of the IPB. Based on this information, a 420 metres deep sample was selected for metagenomic analysis. Finally, a novel microbial species capable of iron biomineralization was isolated from a 297 metres deep sample.

Together, our results indicate the presence of active microbial communities at several depths of the IPB deep subsurface, including those responsible for the genesis of the Tinto River conditions, and provide novel insights on the multiple metabolic strategies used by microorganisms to overcome the challenges posed by deep subsurface environments. Our results show that short-scale geological context determines microbial community composition in deep subsurface ecosystems, with water and oxygen availability being the key factors that select for different microbial metabolisms.

Presentación

La biosfera profunda ha recibido interés creciente a lo largo de las últimas décadas, al ser un ambiente que presenta desafíos únicos para el desarrollo y mantenimiento de la vida. Si bien los primeros estudios mostraron que la abundancia y diversidad de microorganismos era mucho mayor que la esperada originalmente, este ambiente permanece casi completamente inexplorado debido a la dificultad de obtener muestras en condiciones asépticas. Por esto, una serie de preguntas de largo alcance relativas a la energética, filogenia, diversidad metabólica y origen último de estos ecosistemas permanecen todavía sin respuesta.

La Faja Pirítica Ibérica (FPI, suroeste de España) es la anomalía de sulfuros más grande de la corteza terrestre. Desde hace tiempo se hipotetiza que los minerales de su subsuelo son responsables de las condiciones únicas del Río Tinto. Además de esto, el subsuelo de la FPI ha sido reconocido como un análogo de Marte, debido a su parecido geológico con zonas de ese planeta. El estudio de la FPI puede por tanto ayudar a responder varias preguntas clave, incluyendo el grado de influencia de la biosfera del subsuelo en procesos de la superficie, la posibilidad de una vida temprana basada en la litoautotrofía, y el potencial de la vida tal y como la conocemos para desarrollarse en ambientes aparentemente estériles, como Marte en la actualidad.

Durante el desarrollo de esta tesis nos hemos concentrado en la caracterización microbiológica de dos perforaciones (de 160 y 613 metros de profundidad) practicadas en el área de Peña de Hierro, cerca del origen del Río Tinto. Esto ha incluido la optimización de protocolos para aislar y amplificar ADN a partir de muestras minerales, así como el desarrollo de nuevas herramientas bioinformáticas para minimizar la incidencia de contaminaciones y errores de secuenciación en estudios de diversidad por ADN ribosomal 16S. La unión de perfiles taxonómicos con datos fisicoquímicos nos permitió describir los procesos geomicrobiológicos que ocurren en el subsuelo profundo de la FPI. Basándose en esta información, una muestra de 420 metros de profundidad fue seleccionada para análisis metagénomico. Finalmente, una nueva especie microbiana capaz de la biomineralización del hierro fue aislada de una muestra de 297 metros de profundidad.

En conjunto, nuestros resultados indican la presencia de comunidades microbianas activas a varias profundidades del subsuelo profundo de la FPI, incluyendo aquellas responsables de la génesis de las condiciones del Río Tinto, y proporcionan una nueva visión acerca de las múltiples estrategias usadas por los microorganismos para superar los desafíos presentados por estos ambientes. Nuestros resultados también sugieren que el contexto geológico a pequeña escala determina la composición de las comunidades microbianas en el subsuelo profundo, con la disponibilidad de agua y oxígeno siendo los principales factores que seleccionan los distintos metabolismos microbianos.

“If I could make myself look smart just by picking a seemingly profound quote, what would be the point of having written a two hundred pages long thesis?”

Agradecimientos

En primer lugar quiero agradecerle a Víctor Parro el haberme concedido la posibilidad de realizar la tesis doctoral en su laboratorio, así como su constante apoyo a lo largo de estos años. También quiero mostrar mi agradecimiento a mi tutor académico, el Profesor José Berenguer Carlos.

Esta tesis hubiera sido imposible sin la ayuda de muchas personas. Aleyo se adentró conmigo en las profundidades de la tierra. Todavía estamos recorriendo el camino de vuelta. Jacobo aportó su artillería a lo que entonces parecía una guerra perdida. Laurence y Sara me acogieron en el primer mundo, y trataron – sin éxito – de enseñarme algo de química orgánica. Manuel dedicó una hora a explicarme su pipeline de mothur, y cientas a responder las preguntas que, inevitablemente, siguieron. Ninguno de ellos tenía por qué haber movido un dedo por mí. Y sin embargo lo hicieron.

He tenido la inmensa suerte de compartir laboratorio con gente fantástica. Héctor, compañero de fatigas y maestro en la filosofía del desastre. Patri, locura buena en un mundo con demasiados locos malos. Graciela y Nacho, mis hermanos de doctorado. Los anticuerpos nunca fueron (tan) divertidos. Mercedes, Yolanda y Francisco, que tienen sabiduría y paciencia a partes iguales. Luis, que no tiene tanta paciencia, pero sí un don para hacérselo perdonar. Miriam, capaz de hacer que las cosas ocurran.

Quiero también dar las gracias a los compañeros del laboratorio de Extremofilia: Patxi, Ángeles, Sanna, Enoma y Monike, que me han enseñado cómo trabajar con microorganismos en condiciones extremas, y a Mónica y Fernando Tornos, de Geología Planetaria, que me han mostrado que las interacciones entre los minerales y los seres vivos pueden extenderse más allá del uso habilidoso de un tirachinas.

Esta tesis se nutre en gran medida del enorme trabajo realizado por el personal de los proyectos MARTE e IPBSL. En especial quiero agradecer a Marina Postigo por la secuenciación Sanger, David Fernández y Antonio Molina por los datos de geología, Elena González por las fotografías de FISH y Nuria Rodríguez por los datos de gases ocluidos.

Finalmente, quiero dedicar esta tesis a las personas más importantes de mi vida. A mis padres, mi orgullo y mi inspiración. A Irene, mi fuerza y mi armadura.

Abstract

The existence, characteristics and potential limits of the deep subsurface biosphere have been a matter of growing interest over the last decades as it poses unique challenges for the development and maintenance of life. Albeit preliminary studies made clear that the microbial abundance and diversity in the deep subsurface biosphere were much higher than originally anticipated, it remains almost completely unexplored due to the difficulty of obtaining reliable samples. Therefore, a series of wide-ranging research questions regarding the energetics, phylogenetic, metabolic diversity and ultimate origin of such ecosystems are still unanswered.

The Iberian Pyrite Belt (IPB, southwest of Spain) is the most important volcanogenic massive sulphide district in the world and the largest sulphur anomaly in the Earth's crust. It has been long hypothesized that its underground ore bodies are responsible for the unique conditions of the extremely acidic Tinto River, which is born near the sulphide-rich Peña de Hierro mining site. Moreover, the IPB deep subsurface has been recognized as a terrestrial Mars analogue due to its geological similarities to known Martian environments. Its study can thus help to elucidate several key questions, including the degree of influence of deep subsurface life in surfacial events, the possibility of a lithoautotrophic early Earth, or the potential of life as we know it for thriving in apparently barren environments such as the present Mars.

During the course of this thesis we have focused in the microbiological characterization of two boreholes (160 and 613 metres deep) drilled in the Peña de Hierro area, using both conventional and molecular ecology techniques. This has included the development of an optimized protocol for sterile DNA isolation and amplification from mineral samples, as well as the use of novel bioinformatic tools to minimize errors in 16S rRNA gene-based population studies and assess the degree of contamination introduced during sample retrieval and processing. The coupling of microbial taxonomy profiles with physicochemical data allowed us to describe the geomicrobiological processes occurring in the deep subsurface of the IPB. Based on this information, a 420 metres deep sample was selected for metagenomic analysis. Finally, a novel microbial species capable of iron biomineralization was isolated from a 297 metres deep sample.

Together, our results indicate the presence of active microbial communities at several depths of the IPB deep subsurface, including those responsible for the genesis of the Tinto River conditions, and provide novel insights on the multiple metabolic strategies used by microorganisms to overcome the challenges posed by deep subsurface environments. Our results show that short-scale geological context determines microbial community composition in deep subsurface ecosystems, with water and oxygen availability being the key factors that select for different microbial metabolisms.

Presentación

La biosfera profunda ha recibido interés creciente a lo largo de las últimas décadas, al ser un ambiente que presenta desafíos únicos para el desarrollo y mantenimiento de la vida. Si bien los primeros estudios mostraron que la abundancia y diversidad de microorganismos era mucho mayor que la esperada originalmente, este ambiente permanece casi completamente inexplorado debido a la dificultad de obtener muestras en condiciones asépticas. Por esto, una serie de preguntas de largo alcance relativas a la energética, filogenia, diversidad metabólica y origen último de estos ecosistemas permanecen todavía sin respuesta.

La Faja Pirítica Ibérica (FPI, suroeste de España) es la anomalía de sulfuros más grande de la corteza terrestre. Desde hace tiempo se hipotetiza que los minerales de su subsuelo son responsables de las condiciones únicas del Río Tinto. Además de esto, el subsuelo de la FPI ha sido reconocido como un análogo de Marte, debido a su parecido geológico con zonas de ese planeta. El estudio de la FPI puede por tanto ayudar a responder varias preguntas clave, incluyendo el grado de influencia de la biosfera del subsuelo en procesos de la superficie, la posibilidad de una vida temprana basada en la litoautotrofía, y el potencial de la vida tal y como la conocemos para desarrollarse en ambientes aparentemente estériles, como Marte en la actualidad.

Durante el desarrollo de esta tesis nos hemos concentrado en la caracterización microbiológica de dos perforaciones (de 160 y 613 metros de profundidad) practicadas en el área de Peña de Hierro, cerca del origen del Río Tinto. Esto ha incluido la optimización de protocolos para aislar y amplificar ADN a partir de muestras minerales, así como el desarrollo de nuevas herramientas bioinformáticas para minimizar la incidencia de contaminaciones y errores de secuenciación en estudios de diversidad por ADN ribosomal 16S. La unión de perfiles taxonómicos con datos fisicoquímicos nos permitió describir los procesos geomicrobiológicos que ocurren en el subsuelo profundo de la FPI. Basándose en esta información, una muestra de 420 metros de profundidad fue seleccionada para análisis metagénomico. Finalmente, una nueva especie microbiana capaz de la biomineralización del hierro fue aislada de una muestra de 297 metros de profundidad.

En conjunto, nuestros resultados indican la presencia de comunidades microbianas activas a varias profundidades del subsuelo profundo de la FPI, incluyendo aquellas responsables de la génesis de las condiciones del Río Tinto, y proporcionan una nueva visión acerca de las múltiples estrategias usadas por los microorganismos para superar los desafíos presentados por estos ambientes. Nuestros resultados también sugieren que el contexto geológico a pequeña escala determina la composición de las comunidades microbianas en el subsuelo profundo, con la disponibilidad de agua y oxígeno siendo los principales factores que seleccionan los distintos metabolismos microbianos.

Table of Contents

1. INTRODUCTION

1.1. Life in the subsurface

1.1.1. Life in the subsurface: interest, implications and open questions

The existence, characteristics and potential limits of the deep subsurface biosphere have been a matter of growing interest over the last decades (Francis *et al.*, 1989; Gold, 1992; L'Haridon *et al.*, 1995; Lovley & Chapelle, 1995; Chapelle *et al.*, 2002; D'Hondt *et al.*, 2004; Schippers *et al.*, 2005; Kallmeyer *et al.*, 2012). While most surface ecosystems are shaped by photosynthesis as the principal mean of primary production, both terrestrial and oceanic subsurface environments have special characteristics which pose unique challenges for the development and maintenance of life (Fredrickson & Balkwill, 2006). Interestingly, the particular problems that organisms face in those habitats might be similar to the ones encountered in early Earth (Liu *et al.*, 1997; Martin & Russell, 2003; Dong *et al.*, 2014), or those anticipated to occur in extra-terrestrial environments such as Mars or Europa (Boston *et al.*, 1992; Jakosky & Shock, 1998, McCollom, 1999; Fredrickson & Balkwill, 2006; Fernández-Remolar *et al.*, 2005;2008;2008b; Abbey *et al*. 2013). The study of subsurface ecosystems might therefore be capital for elucidating questions of high astrobiological interest, such as the shape and function of the first microbial communities in Earth and the potential of terrestrial-like life for developing in, or colonizing environments that, by our actual standards, are marginally habitable or even plainly inhabitable.

The implications of deep subsurface life are, however, not only limited to the field of astrobiology, but strongly affect ecology and Earth sciences as a whole. During the last centuries the scientific community has come to understand that all the main chemical species required for life, such as carbon, nitrogen, oxygen, phosphorus, sulphur or iron, are not static, but rather cycled through different biotic (the different ecosystems of the biosphere) and abiotic (lithosphere, atmosphere and hydrosphere) compartments (Bolin & Cook, 1983), and that microorganisms are key in ensuring their correct transformation and turnover. Nonetheless, the potential impact of subsurface biology in the global biogeochemical cycles on Earth is still largely unaccounted for.

As more studies were performed, it became evident that the microbial abundance and diversity in such habitats were much higher than originally anticipated. In fact, recent estimations suggest that the majority of the microbial biomass on Earth may reside in the deep subsurface (Whitman *et al.*, 1998; Hinrichs & Inagaki, 2012; Kallmeyer *et al.*, 2012; McMahon & Parnell, 2014), with life

being detected down to four kilometres below land surface (kmbls) (Priscu *et al.*, 1999; Moser *et al.*, 2003) and theoretical calculations extending the potential range of habitability up to six-seven kmbls (Onstott *et al.*, 2014) or even 12 kmbls (Omar *et al.*, 2003; Lin *et al.*, 2006). This increasing body of evidence for an ecologically relevant subsurface biosphere is however hampered by sheer lack of observation: given the extent of the deep biosphere and the difficulties of sampling, the majority of potential habitats remain almost completely unexplored (Edwards *et al.*, 2012). Therefore, a series of wide-ranging research questions regarding the energetics, phylogenetic, metabolic diversity and ultimate origin of deep subsurface ecosystems (Hoehler & Jørgensen, 2013; Teske & Sørensen, 2007; Orsi *et al.*, 2013; L'Haridon *et al.*, 1995) remain unanswered (Wilkins *et al.*, 2014). Thorough and continued research in both terrestrial and marine deep subsurface systems is required if we are to answer those key questions. In order to obtain a whole picture of each ecosystem, the employment of a wide array of techniques for characterizing both the biotic and the abiotic fractions is also essential. Finally, when interpreting the results, we have to bear in mind that the challenging conditions that predominate in the deep subsurface impose severe constraints to the development and sustenance of life. Thus, the appropriate adaptations to such limitations must be searched for and properly characterised in order to satisfactorily describe such unique ecosystems.

1.1.2. Energetics of deep subsurface ecosystems

Subsurface ecosystems are significantly isolated from the exterior, and neither light nor simple photosynthetically-derived substrates are available to sustain life. Therefore, microorganisms have to rely on other organic or inorganic nutrients, such as as buried kerogen, petroleum components, gasses (e.g., H_2 and CH_4) or ions in different states of oxidation (e.g., nitrogen and sulphur species, as well as metals) (Lovley & Chapelle, 1995; Liu *et al.*, 1997; Zhou *et al.*, 2001; Lin *et al.*, 2006; Fredrickson & Balkwill, 2006; Lollar *et al.*, 2006), in order to fuel their metabolisms. Although some nutrients can be recycled within an ecosystem, provided the microbial community is complex enough, at least the main energy source must be provided externally (Morono *et al.*, 2011).

In sub-seafloor sediments, the main source of carbon and energy appears to be the organic matter deposited on the seafloor and buried over millions of years (D'Hondt *et al.*, 2004; D'Hondt *et al.*, 2009; Lipp *et al.*, 2008). In this case, the readily available substrates would be rapidly consumed, leaving only recalcitrant organic matter to sustain microbial growth, which would be increasingly difficult with depth (see Hoehler & Jørgensen, 2013).

An alternative to this scenario that might be particularly relevant in the continental deep subsurface

is a lithoautotrophic biosphere sustained by the products of crustal abiotic reactions (Gold, 1992; Lin *et al.*, 2006; Lollar *et al.*, 2006), which would be able to thrive for millions of years independently of photosynthetically fixed organic matter (Lin *et al.*, 2006). Hydrogen is of special interest, since it can be generated by mineral alteration reactions where mafic and ultramafic minerals dissolve and are replaced by hydrated mineral assemblages (Moody, 1976; Kelley *et al.*, 1994; 2002; Früh-Green *et al.*, 2004; Bach *et al.*, 2006; Schulte *et al.*, 2006; Oze & Sharma, 2007; Bach & Früh-Green, 2010). While its production is often associated to hydrothermal activities, recent studies hypothesize that it might be abiotically produced in the subsurface at relatively low temperatures, with water availability and pore / fracture space as the limiting factors (Hellevang *et al.*, 2011; Okland *et al.*, 2014). Its high molecular diffusivity means that it can easily permeate from the places where it's produced (which will be determined by mineralogy, pore space and water availability) to the consumer microbial communities, while its ubiquity, its ability to penetrate the microbial cell, its low energy of activation, its ability to form protons and electrons in the presence of Fe(II), and its (including electrons and protons) role in many biochemical reactions makes the best candidate as the energy source for microbial cells in the subsurface (Morita, 1999).

Hydrogen is however not the only possible abiotic energy source: sulphur (Ulrich *et al.*, 1998), ferrous iron (Osburn *et al.*, 2014) and methane (Osburn *et al.*, 2014; Teske et al 2014) have been found to be oxidised by microorganisms in different subsurface ecosystems. On the other hand, the range of electron acceptors that can be used for such reactions is broad, and includes CO_2 (Newberry *et al.*, 2004; Chapelle et al 2002), sulphate (Ulrich *et al.*, 1998), ferric iron (Dong *et al.*, 2014), nitrate (Francis *et al.*, 1989; Liebensteiner *et al.*, 2014) or even perchlorate, if available (Liebensteiner *et al.*, 2014). Theoretical calculations have shown that many different redox couples could yield enough energy for sustaining microbial growth, even when taking into consideration the physicochemical conditions of subsurface pore waters (Osburn *et al.*, 2014). The particular processes that predominate in a given ecosystem will depend on the geological, hydrological and physicochemical conditions (Bonte *et al.*, 2013), which are especially variable in the continental subsurface (Osburn *et al.*, 2014).

Even if there is a broad array of potentially energy yielding subsurface metabolisms, it has to be noted that the abiotic reactions required for the generation of energy sources will operate over geologic timescales, that is, in a steady but very slow fashion, especially at low temperatures. In addition to this problem, both microbial activity and geochemical processes are also controlled by the physical properties of the mineral matrix, such as its hydraulic conductivity (Chapelle & McMahon, 1991; McMahon & Chapelle, 1991; McMahon *et al.*, 1992). Lithologies with low pore

space and low water availability will result in limited molecular diffusion and even slower reaction rates for both chemical and biological processes (Jakobsen & Postma, 1994; Fredrickson *et al.*, 1995).

Due to these factors, energy fluxes in deep subsurface ecosystems are thought to be 1000-fold lower than the typical requirements for maintaining basal metabolism (Hoehler & Jørgensen, 2013). In these extremely oligotrophic environments, cells will catabolize 10^4 to 10^6 fold slower and have turn over rates on timescales of centuries to millennia rather than hours to days (D'Hondt et al 2002; Hoehler & Jørgensen, 2013). Thriving under such extreme limitations requires specific adaptations in order to optimize energy conservation and metabolic efficiency over very long periods of time, while maintaining cellular and genetic integrity.

1.1.3. Microbial survivability in very low energy ecosystems

While we are used to think that natural selection will result in organisms that are perfectly adapted to the specific conditions of their environments, that might not be the case in subsurface ecosystem. The slow rates of biomass turnover provide minimal opportunities for the introduction and propagation of beneficial mutations. Thus, organisms in deep subsurface environments will be initially selected according to the traits they had when they arrived to the ecosystem, but they will not be able to further evolve after that (Hoehler & Jørgensen, 2013). Survivability will therefore be determined by traits gained in other ecosystems showing similar – to some extent – restrictions but with higher energy fluxes. In order to understand microbial adaptation to deep subsurface environments, we have to look for examples in which non-subsurface organisms have developed solutions to comparable problems.

Dormancy is a well known strategy for coping with adverse conditions. Endospores can survive over extremely long periods of time with no energy consumption (Lennon & Jones, 2011), and they have been found to be as abundant as vegetative cells in deep, million year old sediments (Lomstein et al 2012). While it is true that dormancy increases the chance that a fraction of the population will survive periods of high environmental stress, this is only useful if the conditions are expected to improve at some point in the future. However, deep subsurface environments are stable over millions of years, which makes dormancy arguably a dead end, since the environmental conditions would never be favourable enough to trigger endospore germination. Moreover, even if some endospores were able to germinate, the additional energetic cost associated with the transition to the vegetative phase would put them in a disadvantage when competing with non-spore formers that are

nonetheless adapted to having very slow metabolic rates.

It has been shown that *E. coli* cells are capable of entering a state known as *extended stationary phase* in which they can persist for months or even years without further addition of substrate (Zambrano *et al.*, 1993; Finkel & Kolter, 1999; Finkel, 2006). This adaptation is based on a dynamic population turnover, in which successive generations recycle the remains of their predecessors with increased efficiency. There are two main factors that make this phenomenon possible: the dead biomass generated during the stationary phase (around 100-1000 dead cell's worth of energy for each survivor of the stationary phase) and a non-limiting supply of an efficient electron acceptor such as oxygen. These conditions will not apply in the low-energy, long-timescales conditions that are characteristic of deep subsurface habitats. Moreover, it has been described that apart from dormant endospores (whose biomass would hardly be available for recycling by other microorganisms) the deep biosphere is mainly inhabited by viable and active cells (Schippers *et al.*, 2005; Hammes *et al.*, 2010). Therefore, the recycling of dead organic mater is unlikely enough for the long-term sustenance of subsurface ecosystems.

Instead of a single metabolic state that acts like a silver bullet adaptation to living in the subsurface, there might be many smaller traits that contribute to an improved metabolic efficiency. The largest single energetic cost that microbial cells will incur in during their life cycles is protein synthesis (Stouthamer, 1973; Stouthamer, 1979). Even if attrition (and therefore the need for repair or replacement) of biomolecules at low rates is unavoidable (Brinton *et al.*, 2002; Lindahl & Karlstrom, 1973) organisms that are able keep the rate of biosynthesis to the minimum needed for basal maintenance will significantly decrease their energy requirements (Hoehler & Jørgensen, 2013). Metabolic regulation involves continuous synthesis and breakdown of proteins, and therefore comes with a very high energetic cost. Moreover, it is unnecessary in deep subsurface environments, which are remarkably stable over geologic times. Therefore, simple and unregulated metabolisms will be the most energy efficient, especially since they would also require smaller genomes which in turn are less expensive to replicate.

Biosynthetic cost is also a key factor, as the same molecules can be synthesized by different pathways with different energy efficiencies. After a rigorous thermodynamic analysis, McCollom & Amend (2005) concluded that the theoretical minimum biosynthetic cost of whole cells is nearly 13-fold lower in reducing versus oxic environments (see also Morowitz, 1968). This involves the usage of less energy intensive pathways, such as the reductive acetyl CoA (Wood-Ljungdahl) pathway rather than the RUBisCO pathway for CO_2 assimilation (Lever, 2011; Schuchmann & Müller, 2014).

Finally, there are some adaptations that minimize energy dissipation during catabolism, increasing metabolic energy yields. It is known that trans-membrane ion gradients, which are required for non-fermentative ATP synthesis, leak over time (Vossenberg *et al.*, 1995). This is especially problematic at the very slow metabolic speeds characteristic of deep subsurface ecosystems. There are two main strategies in order to minimize this loss: use of membranes that are less permeable to ion leakage or use of different ionic species for energy transduction, such as sodium ions (Vossenberg *et al.*, 1995; Vossenberg *et al.*, 2000; Schlegel *et al.*, 2012).

1.1.4. Microbial survivability under extreme water limitations

Microorganisms in the deep subsurface might also face conditions of severe water stress, especially in the vadose (unsaturated) zone above the water table. In fact, water availability has been hypothesized to be the main environmental factor influencing microbial activity in the vadose zone (Kieft *et al.*, 1993). Even when in the saturated zone, low permeability lithologies will put limits on microbial metabolic rates (Fredrickson *et al.*, 1995). Desiccation is fatal to the majority of organisms, since it results in alterations to all cellular components: biological membranes experiment phase transitions that may result in leakage during rehydration, proteins undergo conformational changes and DNA is substantially damaged by strand breaks, base modifications, abasic sites and cross links (Dose *et al.*, 1992;2001; Potts, 1994; Billi & Potts 2002). Some of these damages are oxidative, and their effect might be mitigated in reducing environments, such as in the presence of sulphide (Potts *et al.*, 2005). However, microorganisms still need special adaptations in order to cope with such extreme conditions.

Gram-positive bacteria present several adaptations that makes them generally more resistant to desiccation (Potts *et al.*, 2005), including the accumulation of Mn^{2+}, a high Mn^{2+}/Fe^{2+} ratio, and radiation tolerance mechanisms (Daly *et al.*, 2004). Radiation resistance, in particular, has been correlated with desiccation tolerance (Slade & Radman, 2011), which suggests that both processes could be molecularly regulated by similar mechanisms. On the other hand, cyanobacteria, despite having a Gram-negative wall, show phylogenetic affinity to Gram-positives and are significantly resistant to both desiccation and gamma radiation (Potts, 1999; Billi *et al.*, 2000).

In general terms, the loss of water during the desiccation process is balanced due to reversible interactions with other molecules (Rebecchi *et al.*, 2007), preventing molecular degradation and allowing biomolecules and structures to retain their native conformation after rehydration (Wolkers *et al.*, 2002). Intracellular accumulation Mn^{2+} has been proposed as the most probable mechanism

involved in the protection of proteins from oxidative damage after desiccation (Fredrickson *et al.*, 2008). The accumulation of non-reducing disaccharides such as sucrose and trehalose (Clegg, 2001; Rebecchi *et al.*, 2007) is another strategy for desiccation tolerance. These sugars are able to replace the water in biomolecules by forming hydrogen bonds, and contribute to the formation of a vitreous cytoplasmic matrix (Wolkers *et al.*, 2002). Stress proteins seem to be also involved in anhydrobiosis, especially Heat Shock Proteins (HSP) (Goyal *et al.*, 2005; França *et al.*, 2007), late embryogenesis abundant (LEA) (Rebecchi *et al.*, 2007), and oxidative stress response proteins such as catalase and superoxide dismutase (Markillie *et al.*, 1999), which are found even in some strictly anaerobic microorganisms (de Macario & Macario, 2000).

In addition to resisting the adverse effects of desiccation, microorganisms need to be able to obtain and preserve a minimum amount of water in order to maintain metabolic activity. Several possible sources of liquid water in otherwise dry environments have been proposed, such as interfacial water present in layers of only one to several molecules forming on mineral surfaces by adsorption (Dash *et al.*, 2006; Möhlmann, 2011) or brines formed due to deliquescence on salt crystals. Microbes can also obtain water from the gas phase, a process that has been observed in lichens (Pintado & Sancho, 2002; Lange *et al.*, 2006). Metabolic activity can be another important supply of water, and radio-labelled gas uptake experiments have suggested that up to 70% of the intracellular water might have a metabolic origin (Kreuzer-Martin *et al.*, 2005;2006). Several evidences have led to the hypothesis that microorganisms may be capable of dividing without an extracellular supply of liquid water (Miller & Chibnall, 1932; Yarwood, 1950; Peterson & Cowling, 1973; Lange et al 1986;1994). However, this is not exempt of controversy, as systematic studies in several extremophilic microbes suggest that cell division would be implausible at water activities below 0.6 (Pitt & Christian, 1968; Brown, 1976; Williams & Hallsworth, 2009).

1.1.5. Depth limits of the subsurface biosphere

Microorganisms have to overcome several challenges in order to thrive in deep subsurface environments, which become increasingly hostile as depth increases. Eventually, there will be a boundary, determined by the geological characteristics of the system, below which life will not be able to develop.

It is currently accepted that the upper temperature limit for life as we know it is approximately 121ºC (Kashefi & Lovley, 2003), although temperatures up to 150ºC have been considered (Deming & Baross, 1993). Geothermal gradients in the continental crust are variable, depending on the

ability of the geological material to conduct heat, but on average there is a temperature increase of 25ºC per kilometre (Akob & Küsel, 2011; Fredrickson & Fletcher 2001; Pedersen, 2000). This gives us a depth limit of approximately 5 kmbls (Akob & Küsel, 2011), which might extend down to 12 kmbls in locations with less thermally conductive mineralogies (Lin *et al.*, 2006; Omar *et al.*, 2003).

These figures can be further elaborated if we consider the actual effects of temperature in microbial survivability. Maintenance energy exponentially increases with temperature (Roels, 1983; Tijhuis *et al.*, 1993; Röling *et al.*, 2003), due to the need of rapid anabolic rates in order to replace damaged cellular constituents. In fact, the amount of energy required to repair protein and DNA damage have been proposed to constrain the minimum energy requirement for microbial survival (Morita, 1997; Hoehler 2004), and the rates for theses processes are strongly temperature dependent (Price & Sowers, 2004). Therefore, even for hyperthermophilic strains, the actual ability to grow at high temperatures will largely depend on energy availability, with the maximum of 121ºC being only reachable in high energy flux environments such as thermal springs and deep-sea hydrothermal vents. Those favourable conditions are unlikely to occur in the deep subsurface, except in exceptional cases like in the presence of anomalous radiolytic hydrogen fluxes (Lin *et al.*, 2005). This places an additional constraint to the depth limit of the subsurface biosphere.

Hydrostatic pressure, which also increases with depth, can also put a lower limit to the deep biosphere. Microorganisms can generally tolerate pressures of 10 to 100 MPa, with some barophilic organisms being able to withstand pressures of up to 1000 Mpa (Fredrickson & Onstott, 1996; Sharma *et al.*, 2002). In the continental crust, the hydrostatic pressure at a given depth will depend on the depth to the groundwater within the system. It is generally believed that temperature is more likely to limit microbial growth before elevated hydrostatic pressures are reached (Fredrickson & Onstott, 1996); however, this might not be the case in subduction zones, in which there are smaller temperature gradients (Colwell & Smith, 2004).

Finally, lithology will influence the lower boundary of the subsurface biosphere not only because of the effect of different thermal conductivities, but also because it will determine the available pore space, which is required for colonization and nutrient diffusion (Colwell *et al.*, 1997; Colwell & Smith, 2004). The initial porosity of sediments will be different depending on the material (Poelchau *et al.*, 1997), and these differences will further increase with depth as the rock compacts with pressure, diagenesis and tectonic stress (Chilingarian, 1983). Permeability, which is the ability of sediments to transmit fluids and/or gasses, is controlled by the effective porosity and the number and type of pore interconnections. Volcanogenic sandstones will quickly lose porosity and

permeability with depth, while in contrast quartz arenites offer perhaps the most stable living accommodation for microorganisms (Horsfield & Kieft, 2007). Fractures will provide additional space and communicate aquifers with different compositions, generating chemical gradients that can be used by microorganisms to produce energy (Pedersen, 2014). Thus, the relationship between thermal conductivity (which, as discussed above, determines the lower energy limit for microbial survivability at a given depth) and permeability (which greatly affects energy availability) might be key in determining the depth limits of microbial life.

1.2. Classification of deep subsurface environments

1.2.1. Marine sediments

Marine sediments result from the accumulation of particles that sink to the sea floor from the overlying water column. The whole seafloor is covered by a layer of sediments, although its thickness varies from almost nonexistent (near newly formed crust or beneath the low productivity zones of the ocean gyres) to deposits of about 10 km thick (at trenches and highly productive continental margins). Their composition is determined by the origin of the buried material, and can include biological material (mainly diatom shells that deposit from the surface ocean), eroded terrestrial matter delivered by rivers, land particles deposited by the wind or hydrothermal-vent delivered minerals. Vertically, sediments are classified as either shallow (from a few centimetres to metres of depth) or deep.

Early research focused on the superficial sediments from the continental margins (ZoBell & Anderson, 1936; ZoBell 1938; ZoBell & Morita, 1957), mainly due to their higher accessibility, but also because they play a key role in the remineralization of particulate organic matter (Burdige, 2006). These shallow sediments are generally characterised by a high availability of organic matter, and a limitation of electron acceptors (mainly oxygen, nitrate and sulphate but also manganese, iron and carbonate), whose availability will determine the degree of organic matter remineralization. This results in steep geochemical gradients and high local rates of microbial activity. The readily degradable substrates are consumed in the first centimetres of sediments, but a large part of the organic matter, especially the most recalcitrant compounds, will escape oxidation and thus be available to support microbial growth in the deeper zones of the sediment (Middelburg, 1989; Parkes *et al.*, 2000).

In spite of the aforementioned sampling difficulty, deep subsurface marine sediments have been

increasingly studied since the late 1980s (Cragg *et al.*, 1990; Parkes *et al.*, 1994; D'Hondt *et al.*, 2004; Hinrichs et al 2006; Biddle, 2006; Inagaki *et al.*, 2006; Engelen *et al.*, 2008; Lipp *et al.*, 2008). As described in previous sections, these ecosystems are thought to be energy limited, but in the case of deep marine sediments there is a constant, albeit slow, influx of organic matter from the above. This suggest that degradation of buried organic matter may be the main energy source for microbes in such environments, and a positive correlation between organic carbon content and microbial community size has been confirmed in various studies (D'Hondt *et al.*, 2002; Lipp *et al.*, 2008; Parkes *et al.*, 2005). At higher depths, the increase in temperature might cause modifications in the remaining refractory compounds, increasing their bioavailability (Wellsbury *et al.*, 1997). In the deepest layer of the sediment, at the interface between sediment and the underlying ocean crust, the oxidised crustal fluids provide a new input of electron acceptors, possibly stimulating microbial activity (D'Hondt *et al.*, 2004; Parkes *et al.*, 2005).

1.2.2. Marine crust

Below the marine sediment layer lies the igneous ocean crust. Little is known about its biosphere, since as of today it remains poorly studied (Schrenk *et al.*, 2010). However, in the last decade several works have preliminary characterised these ecosystems, suggesting that microbial life in such environments is present over a range of temperatures, crustal ages, fluid compositions and redox conditions (Fisk *et al.*, 1998; Cowen *et al.*, 2003; Santelli *et al.*, 2008; Staudigel *et al.*, 2008; Mason *et al.*, 2010)

The igneous ocean crust is mainly comprised of an upper layer of volcanic basalts followed by a deeper one of plutonic gabbros, and hosts the largest aquifer on Earth, representing approximately 2% of the ocean fluid volume (Johnson & Pruis, 2003). The igneous ocean crust is hydrothermally active, and there is an exchange of fluids between the crust and the overlying oceans, especially at ridge flanks where the sediment layer is thin or nonexistent (Wheat *et al.*, 2003). The high hydrological activity facilitates nutrient diffusion and microbial colonization (Edwards *et al.*, 2011), resulting in an environment with a much higher potential habitability than the sediments above it (Orcutt *et al.*, 2011). In addition to this, the regions with hydrothermal influence (which have been the most extensively studied) are enriched in reduced compounds, such as hydrocarbons, sulphide, hydrogen or metals, that can be oxidised by thermophilic microbes in order to obtain energy (Bach & Edwards, 2003). Even in the absence of direct hydrothermal activity, basalt alteration through iron oxidation is thought to be able to provide enough energy to sustain mesophilic ecosystems.

Studies in superficial basalts have revealed surprisingly diverse populations (Mason *et al.*, 2008; Santelli *et al.*, 2008;2009; Orcutt *et al.*, 2011), especially in the older, more altered basalts (Santelli *et al.*, 2008;2009). While it is unclear if those results can be extended to the deeper gabbroic layer, the high permeability and porosity of the igneous ocean crust suggests that this may be the case. In what, to the best of our knowledge, constitutes the only report in the microbiology of the gabbroic layer, Mason *et al.*, 2010 reported the presence of a hydrocarbon-driven ecosystem, down to a depth of 1391 metres below the sea floor. Interestingly, they found a similar correlation between microbial diversity and rock alteration as the one described by Santelli et al (2009) for superficial basalts. The potentially high energy fluxes provided by hydrothermal activity near the mantle could allow for the maintenance of ultra-deep thermophilic microbial communities, as described in the previous section. However, as in other subsurface environments, the scarcity of data due to the extreme sampling difficulty is a huge barrier for the confirmation of any proposed hypothesis.

1.2.3. Continental crust

In opposition to the relative homogeneity of the marine crust, the continental crust is much more geologically diverse. It is formed by volcanic, sedimentary, and metamorphic rocks, and it is older, less dense, and generally more felsic than its oceanic counterpart. It is also much thicker, ranging from 25 to 75 km in opposition to the average 7-10 km thickness of the oceanic crusts. This means that, even if only 40% of the Earth's surface is covered by continental crust, it makes up for about 70% of the total crustal volume.

The volume and heterogeneity of the continental crust make a global description of its physicochemical conditions and associated biosphere a very challenging task – perhaps even an impossible one (Edwards *et al.*, 2012). Its ecology has been traditionally studied by analysing fracture water samples and their associated biofilms, as fracture systems are thought to contain the majority of the continental deep subsurface biosphere (Onstott *et al.*, 2003). A wide array of heterotrophic and lithoautotrophic metabolisms have been detected, including fermenters, sulphate and iron reducers, nitrogen fixers, methanogens or acetogens (Slater *et al.*, 2006; Chivian *et al.*, 2008; Davidson *et al.*, 2011; Itävaara *et al.*, 2011; Osburn *et al.*, 2014; Dong *et al.*, 2014; Purkamo *et al.*, 2014; Blanco *et al.*, 2014), with the overall heterogeneity observed between studies and sample types suggesting a site-specific microbiota controlled by the mineral microenvironment (Jones & Bennet, 2014). Microbial abundances are highly variable but generally decrease with depth (Onstott et al., 1998; Whitman *et al.*, 1998; Pedersen, 2000; Fang & Zhang, 2011; Itävaara *et*

al., 2011). Most strikingly, at least one single species ecosystem has been described in the crustal deep subsurface (Chivian *et al.*, 2008), which has tremendous implications in the fields of astrobiology and early life research.

Direct sampling of subsurface rock cores have been far less common – due, again, to the extreme difficulty of aseptic sample retrieval. Recent efforts have confirmed the presence of deep subsurface microorganisms living within natural pore spaces in crystalline rocks. However, and to the best of our knowledge, these works have been performed either in ultra deep gold mines (Colwell *et al.*, 1997; Onstott *et al.*, 2003; Rastogi *et al.*, 2010; Blanco *et al.*, 2014) or in pre-existent oil surveys (Onstott et al., 1998). This greatly facilitates sample recovery, but compromises its microbiological integrity, as human activity in both environments might have introduced an unknown degree of contamination (Blanco *et al.*, 2014). Regardless of its cost and complexity, a comprehensive set of dedicated scientific drilling campaigns performed in different geological settings would be required in order to truly elucidate the extent, composition and global biogeochemical relevance of the continental deep subsurface biosphere (Fang & Zhang, 2011).

1.3. The Iberian Pyrite Belt as a case study

1.3.1. The Iberian Pyrite Belt and the Tinto River origin

The Iberian Pyrite Belt (IPB) is located in the south west of the Iberian peninsula, comprising part of Portugal and of the provinces of Huelva and Seville in Spain. It is the most important volcanogenic massive sulphide district in the world and the largest sulphur anomaly in the Earth's crust (Tornos, 2006), having been mined during more than 5000 years.

Stratigraphically, it is composed by a basal unit with more than 2000 metres of slate and sandstone (Phyllite-Quartzite Group or PQ Group), overlain by a Volcano-Sedimentary Complex (VSC) that reaches a thickness of 1300 metres. This complex was deposited in an intracontinental marine basin during the Late Devonian – Early Carboniferous, and it is mainly felsic, showing compositions from basalt to rhyolite. Finally, the Culm Group diachronically lays on the VSC and consists of a synorogenic flysch with an Early Carboniferous age. The whole series has been affected by very low degree metamorphism and fold – thrust tectonics (Silva *et al.*, 1990; Quesada, 1996).

Most of the mineral deposits in this area are located within the Volcano-Sedimentary Complex (Leistel *et al.*, 1998; Carvalho *et al.*, 1999), and were generated during an important seafloor hydrothermal activity that deposited stratiform massive sulphides, mainly composed by pyrite but

also sphalerite, galena and chalchopyrite. Beneath these lenses there is a network of sulphide-rich veins considered as the zone that channelled hydrothermal fluids on their way to the surface (stockwork zone). These sulphides are believed to have originated from the circulation of deep fluids equilibrated with the PQ Group, which would also be the source of the metals and part of the sulphur (Tornos, 2006). The massive sulphides would have precipitated by the mixing of these fluids with sea water, with most of the sulphate being biologically reduced to sulphur.

Even if the IPB is no longer experiencing present day hydrothermal events, the past activity has left behind vast amounts of diverse chemical species that could potentially sustain a complex deep subsurface ecosystem. The oxidation and erosion, both by human and natural means, of the massive sulphides significantly altered the IPB environment. Water draining these ore bodies is characterised by an extremely acid pH and thus carries important amounts of metals, resulting in rivers with unique ecosystems, of which the Tinto River is the paradigmatic example.

The basin of the Tinto River covers an area of 1676 km^2 in the province of Huelva (southwest region of Spain). From its source at Peña de Hierro (altitude 500 metres), it has a course of 92 km until reaching the Atlantic Ocean in Huelva. Its slope is gentle, facilitating the settlement of a dense microbial community on the riverbed. The river flow is extremely variable, due to the seasonal changes in temperature and precipitation that are characteristic of a semi-arid Mediterranean-type climate. However, the pH remains low (mean value of 2.2) and constant, due to the buffer effect produced by high concentrations of ferric iron along the river (González-Toril *et al.*, 2003). This extremely low pH, together with the high concentration of sulphate, ferric iron and other heavy metals, make the Tinto River very similar to acidic mine drainage systems (AMD) (Johnson, 2001; Hallberg & Johnson, 2001). These environments are characterised by the outflow of acidic waters from metal mines or coal mines, due to the weathering of sulphides exposed to water and oxygen by mining activities. Pyrite oxidation is mainly catalysed by acidophilic microorganisms such as *Acidithiobacillus, Leptospirillum* or *Acidiphilium,* which are indeed relevant in the Tinto River (López-Archilla *et al.*, 2001; González-Toril *et al.*, 2003). Even so, and unlike other AMD environments, the Tinto River hosts a surprisingly complex microbial community, in which acidophilic photosynthetic eukaryotes are the main contributors to biomass (Amaral-Zettler *et al.*, 2002; Aguilera *et al.*, 2006). Additionally the suboxic and anoxic sediments of the Tinto River are highly diverse, and play a significant role in iron and sulphur cycling (García-Moyano *et al.*, 2012). Together, this features result in a unique ecosystem of high astrobiological interest (Amils *et al.*, 2014).

For decades, the extreme conditions of the Tinto River were thought to be the direct consequence of the intensive mining activity carried out in the area (Van Geen *et al.*, 1997; Elbaz-Poulichet & Dupuy, 1999; Davis *et al.*, 2000; Leblanc *et al.*, 2000). However, recent evidences point to a natural origin of the Tinto River. The sedimentary record shows evidences of massive laminated iron bioformations similar as the ones being currently formed in the river, but dating from 2.1 million years ago (Fernández-Remolar *et al.*, 2003;2005). This suggests that an ecosystem similar as today's was in place way before the beginning of human activity in the area. Furthermore, the location of the recharge area of the Peña de Hierro aquifer has been determined to be at a depth ranging from -100 to -400 mbsf, and water is thought to cross the massive sulphide bodies before reaching the surface, driving microbial pyrite oxidation and being acidified in the process (Gómez-Ortiz *et al.*, 2014). This interpretation is not yet uncontested (see Olías & Nieto, 2014), but it is especially interesting. If confirmed, the presence of a continuous underground reactor responsible for the extreme conditions of the Tinto River would be an excellent example of how the deep subsurface biosphere can exert a decisive influence in surface ecosystems.

1.3.2. The Iberian Pyrite Belt subsurface and the iron sulphur world hypothesis

Until recently, evolution theory has been dominated by the notion of an heterotrophic origin of life (Oparin, 1924; Bernal, 1949), albeit a few authors proposed an autotrophic origin (Hartman, 1975; Woese, 1979). Wächtershäuser (1990), in particular, introduced the idea that pyrite formation from ferrous sulphide, which was ubiquitous in the hydrothermally active Archean Eon, would have been a perfect candidate for single-step carbon dioxide fixation, releasing energy in the process as described in the equation:

$$4CO_2 + 7FeS + 7H_2S \rightarrow (CH_2COOH)_2 + 7FeS_2 + 4H_2O \quad \Delta G^\circ = -420 \text{ kJ/mol}$$

Pyrite is also able to bond anions, such as the carboxylates generated during carbon fixation, due to its positively charged surface. This means that the products of the aforementioned reaction would not diffuse to water, but accumulate in the mineral surface, being available for further reactions that may evolve into a primitive surfacial proto-metabolism (Wächtershäuser, 1988,1990) and, eventually, cellularization. This led to the establishment of the iron-sulphur world hypothesis for the origin and early evolution of life (Wächtershäuser, 1990; see also a critique of Wächtershäuser's model in de Duve & Miller, 1991) of which some key points have been proven experimentally (Cody *et al.*, 2000; see also Wächtershäuser, 2000 for further analysis).

Regardless of its ultimate origin, it is universally accepted that early life had to cope with an environment very different from today's. During the Archaean Eon, before the uprising of cyanobacteria, free oxygen was almost absent and, as a consequence, iron concentrations may have been as high as 50 μM (Holland, 1973), three orders of magnitude higher than in today's oxygenated seawater (Anbar & Knoll, 2002). The high iron availability, together with the fact that ferric iron respiration has been postulated to be a very primitive mechanism for energy conservation (Lovley, 2000), suggests an Archaean biosphere dominated by a fully operative microbial iron cycle (Gómez & Amils, 2002; Amils *et al.*, 2005). In fact, this biosphere would not even require high concentrations of dissolved iron, as chemolithotrophic microorganisms can efficiently obtain energy from insoluble iron minerals (Lovley, 2000; Blake & Johnson, 2000).

Earth's subsurface has been proposed as a favourable location for the origin of life, since it could have provided both protection from radiation and nutrients (Trevors, 2002). The Iberian Pyrite Belt subsurface, rich in both iron and sulphides, shows striking similarities to some ancient Earth environments, making it a very useful proxy for testing hypotheses concerning the origin and early development of life.

1.3.3. The Iberian Pyrite Belt subsurface as a Mars analogue

The discovery of sulphate deposits on Mars (Gendrin *et al.*, 2005; Squyres *et al.*, 2004;2006) and the predominance of iron minerals on its composition has shed light into the role of iron and sulphur geochemistry as essential processes for understanding the potential past and present habitability of said planet. The present lack of an atmosphere, which would have provided greenhouse warming and protection from radiation (Fanale *et al.*, 1992; Squyres & Kasting, 1994; Brain & Jakosky, 1998), makes the subsurface the most obvious region in which to search for potential Martian habitats (McKay, 2001).

Additionally, the recent detection of jarosite, a ferric iron sulphate-hidroxide mineral, on the Martian surface at Meridiani Planum (Klingelhöfer *et al.*, 2004) placed a particular constraint in the interpretation of Mars paleo-environment, as said mineral is considered to precipitate only in acidic conditions (Bigham *et al.*, 1996). Its presence requires a wet, oxidizing and acidic environment in early Mars, albeit its persistence over time suggests that, following jarosite formation, arid conditions prevailed (Madden *et al.*, 2004). Jarosite, as well as other characteristic minerals such as goethite and hematite, are abundant in the Tinto basin as a result of the activity of chemolithotrophic microorganisms that thrive in the iron sulphides of the IPB (Fernández-Remolar

et al., 2005). This makes the Tinto River basin a very useful Mars analogue, not only as a proxy of its past conditions, but also for hypothesizing about extant processes (see Amils *et al.*, 2014 for a more thorough review).

The recent detection of methane in Mars' atmosphere led to several hypotheses about its sources, including a biogenic generation by a past or present subsurface biosphere (Formisano *et al.*, 2004; Mumma *et al.*, 2009). As discussed above, the subsurface is indeed the most likely (or at least the least unlikely) place to host a hypothetical Martian biosphere, but the question remains of whether a geologically dead planet can support extant life. Nonetheless, the mineral deposits generated during hydrothermal activity can last over hundreds of millions of years, as it is the case with sulphides in the IPB subsurface (see **section 1.3.1**), being potentially able to maintain substantial energy fluxes even when said activity is long gone. Due to its mineralogical similarities to Mars and the presence of plausible energy sources independent from light or present geologic activity, the IPB subsurface is the optimal place to test hypotheses on Martian habitability. Previous works have proposed that a hypothetical Martian biosphere thriving in a subsurface environment similar to the one present in the IPB would be able to maintain pH and thermal homoeostasis by negative feedback (Fernández-Remolar *et al.*, 2008b), not only taking advantage of pre-existent habitable zones, but actually increasing their own environment's habitability.

As it has been discussed in this and previous sections, the IPB subsurface holds special relevance for answering several key questions, including the degree of influence of deep subsurface life in surfacial events, the possibility of a lithoautotrophic early Earth, or the potential of life as we know it for thriving in apparently barren environments such as the present Mars. However, little is known about the actual microbiota that inhabits it. Understanding the particular biological processes that operate in such a unique environment will greatly expand our knowledge on the potential and limits of life in this and other planets.

1.4. Techniques for the study of the deep subsurface biosphere

1.4.1. Challenges in aseptic sample retrieval

Recovering subsurface samples generally requires complex drilling technologies, although in some cases pre-existing infrastructures, such as gold mines, may be used. Several drilling techniques are available for shallow sampling of the continental subsurface (see Kieft, 2010), but while they can be used without drilling fluids (thus limiting the potential for contamination), rotary drilling with

addition of drilling fluids is required for accessing materials deeper than 300 mbsf (Wilkins *et al.*, 2014). These fluids contribute to seal the borehole and cool and lubricate the drilling bit, but they often carry high loads of microorganisms. Besides exerting an utmost care during sample retrieval, the addition of chemical or biological tracers to the drilling fluids is essential in order to keep track of the potential contaminations (Beeman & Suflita, 1989, see Wilkins *et al.*, 2014 for a more thorough review).

Once a core has been retrieved, it must be immediately subsampled prior to geochemical and microbiological analyses. As for core retrieval, an aseptic (and anoxic, if required) subsampling must be ensured. There are several techniques available for subsampling, including core extrusion, hammer and chisel, hydraulic crushing or crushing with mortar and pestle (see references in Wilkins *et al.*, 2014) and the optimal choice will ultimately depend on the lithology and intended subsequent analyses. Finally, subsamples must be immediately preserved with appropriate methods, normally rapid-freezing for samples destined to biomarker studies, or mild refrigeration for samples destined for culturing experiments (Haldeman *et al.*, 1994).

1.4.2. Physicochemical characterisation of deep subsurface samples

As discussed in the previous sections, microbial communities are interrelated with their habitat's physicochemical parameters, as both will influence each other. Linking community composition with the presence of electron donors and acceptors can help in inferring the most likely metabolisms that are present in the ecosystem. Ion chromatography can be used to determine the presence of anions such as sulphate, nitrate, nitrite or small organic acids in pore water or rock leachates (Fredrickson *et al.*, 1995; Fernández-Remolar *et al.*, 2008; Amils *et al.*, 2013; Blanco *et al.*, 2014) while inductively coupled plasma mass spectrometry (ICP-MS) can in turn determine the presence of metallic and non-metallic elements in mineral samples (Morgan 2005; Amils *et al.*, 2013). The abundance and evolution of gasses like carbon dioxide, hydrogen or methane in pore fluids can be measured by gas chromatography (Fredrickson *et al.*, 1995; Fernández-Remolar *et al.*, 2008; D'Hondt *et al.*, 2009; Amils *et al.*, 2013).

Microbial activity is not only determined by the chemical environment, but also by rock mineralogy and lithology (see **section 1.1.5**). A significant understanding of the mineralogy and geochemistry of a subsurface environment can be derived from the effective utilization of well log data (Onstott *et al.*, 1998; Wilkins *et al.*, 2014). Finally, imaging techniques such as scanning electron microscopy coupled with energy-dispersive X-ray spectroscopy (SEM-EDX) can provide details on

rock structure and composition at the micrometric level, revealing potential physical and chemical micro-habitats (Fernández-Remolar *et al.*, 2008; Tornos *et al.*, 2014).

1.4.3. Organic geochemistry

Living organisms are the main source of complex organic compounds on Earth, and many of these geochemical signatures can be preserved over geologic time. The isolation and characterisation of such organic compounds can therefore provide great insights into the composition, structure and activities of past and present deep subsurface microbial communities (Wilkins *et al.*, 2014).

Lipid biomarker profiles can be determined from intact polar lipids or their derived fatty acid methyl esters to provide estimates of biomass and determine the relative abundance of broad taxonomic groups, including Eukaryotes, Bacteria, and Archaea (Fredrickson *et al.*, 1995; White *et al.*, 1996). On the other hand, polar lipid analyses can be used to estimate total biomass and the proportion of viable versus dead cells (Balkwill *et al.*, 1988; Findlay *et al.*, 1989; Fredrickson *et al.*, 1995). Changes in the relative proportion of fatty acids have also been used to infer metabolic states, such as stress (Heipieper *et al.*, 1992; White & Ringelberg, 1997), or starvation (Guckert *et al.*, 1986; Kieft *et al.*, 1994), while respiratory quinones have been employed to infer environmental redox potentials (White & Ringelberg, 1997).

Finally, biological processes often involve large isotopic fractionation, and are actually the most important cause of variations in the isotope composition of carbon, nitrogen, and sulphur (Hoefs, 2008). Different metabolic pathways will leave distinct isotopic signatures in the resulting products, thus keeping a record of the metabolic activities that generated them.

1.4.4. Microscopy techniques

Direct light microscopic observation, combined with specific staining methods – including the use of fluorescent stains such as 4',6-diamidino-2-phenylindole (Porter & Feig, 1980) – is often used for estimating the microbial content in deep subsurface environments. However, it must be noted that the mineral matrix may difficult the observation of microbes by fluorescence microscopy, as many minerals are themselves autofluorescent.

Fluorescence *in situ* hybridization (FISH, Langer-Safer *et al.*, 1982) and improved variants such as catalysed reporter deposition fluorescence *in situ* hybridization (CARD-FISH, Schönhuber *et al.*,

1997) allow for the microscopic detection of specific DNA and RNA sequences, even in mineral samples. Different hybridization probes can be designed, targeting from broad microbial taxa to single species, with the advantage that only living microorganisms with intact nucleic acids will be reported. This powerful and versatile technique has seen major use in many molecular ecology studies, including those performed in deep subsurface environments (Webster *et al.*, 2009; Schippers *et al.*, 2005; Schippers *et al.*, 2012).

Finally, SEM-EDX can be used to search for microbial structures in mineral samples and is useful to determine past or present microbe-mineral interactions (Folk, 1993; Ventura *et al.*, 2007; Fernández-Remolar *et al.*, 2008). Caution must however be exerted when interpreting SEM images on complex samples, since since some abiotic processes can produce bacteria-shaped microstructures (Bradley *et al.*, 1997).

1.4.5. Culture-based approaches

The use of microbiological cultures has been capital in microbial ecology since its foundation by Beijerink and Winogradsky at the end of the nineteenth century. The isolation of the microorganisms present in a given environment has long been the only way to elucidate their relationships with one another and with their environment, and have therefore been applied in virtually every accessible ecosystem, including deep subsurface habitats (Cragg *et al.*, 1990; Kieft *et al.*, 1993; Parkes *et al.*, 1994; Onstott *et al.*, 1998; D'Hondt *et al.*, 2002; D'Hondt *et al.*, 2004).

The development of enrichment cultures, in which the culture medium is designed to select for microorganisms with specific metabolic traits, allowed microbiologists to explore the physiological and metabolic potential of microbial life, and has led to the present understanding of the key role of microbes in the global biogeochemical cycles. More advanced techniques, such as the use of mixed cultures for elucidating syntrophic relationships between microbial species, or microcosms for studying the assembly and maintenance of complex microbial ecosystems under controlled conditions, are also widely applied. A full review can be found in Wolfaardt *et al.*, 2007. Despite the arrival of new techniques to assess microbial community composition, structure and metabolic potential (see **section 1.4.6**), the power, versatility and conceptual simplicity of culture-based approaches have made them remain a valid tool for modern environmental microbiology studies.

1.4.6. The 'omics' in the study of subsurface life

1.4.6.1. Metagenomics

Culture-based approaches are useful in their own right, but rely on the ability of microorganisms to grow in artificial culture media. It is now common knowledge that this is not the case for the vast majority of the microbial diversity, which was therefore virtually ignored until the introduction of molecular ecology techniques (Hugenholtz *et al.*, 1998). Early work focused in the amplification, cloning and sequencing of the 16S rRNA gene, which is highly conserved but has enough differences between species to be a reliable taxonomic marker (Lane *et al.*, 1985). The generation of 16S profiles for assessing the microbial community composition of a sample is now a widespread technique, and has led to the discovery of many novel taxa and provided novel insights into previously uncharacterised ecosystems (Hugenholtz, 2002). In some cases, microarrays containing probes against a comprehensive set of microbial taxa have been used instead of sequencing for 16S community profiling (Garrido *et al.*, 2008; Rastogi *et al.*, 2010)

The potential of environmental DNA sequencing is not limited to the study of a single marker gene, as it can also aim to sequence – even if only partially – all the genomes (that is, the metagenome) of a microbial community (Healy *et al.*, 1995; Stein *et al.*, 1996). This way not only taxonomic composition can be determined, but also the metabolic processes that are possible in the community (Hugenholtz, 2002). The arrival of high-throughput sequencing technologies in the mid 2000s exponentially increased the amount of information that could be obtained in a single sequencing run, while dramatically lowering the sequencing cost per base, a tendency that has continued ever since (Mardis, 2008). The decreased cost and increased flow of information has led to a general adoption of metagenomics for the study of microbial communities (see Segata, 2013 for a full review), including the ones present in deep subsurface environments (Mason *et al.*, 2008; Chivian *et al.* 2008; Osburn *et al.*, 2014).

1.4.6.2. Metatranscriptomics

Metagenomics can highlight key functions or metabolic pathways that could be potentially determinant for understanding microbial communities and their interaction with their environment. However, it can not determine the actual functions that are active at a specific time and place, or how those activities change in response to environmental forces or biotic interactions. In order to obtain that information, the metatranscriptome of the community must be studied by isolating and sequencing its mRNA.

Albeit metatranscriptomics is conceptually similar to metagenomics, several practical considerations make it much harder to apply, such as the difficulty of isolating intact RNA from complex samples and separate bacterial mRNA from the less informative rRNA. The lack of a predictable relationship between mRNA abundance and protein activity has been also pointed out as a theoretical problem (Moran, 2010).

Nevertheless, solutions to these practical problems have been developed over time (Stewart *et al.*, 2010) and thus this technique has been applied to several microbial communities in the last years (see Stewart et al 2012; Turner *et al.*, 2013). In the case of deep subsurface environments, however, isolating a sufficient amount of intact RNA poses additional challenges (see next sections). To the best of our knowledge, only one metatranscriptomic study of a deep subsurface ecosystem has been published so far (Orsi *et al.*, 2013).

1.4.6.3. Nucleic acids extraction and amplification

In order to successfully apply sequencing techniques, sizeable amounts of high-quality nucleic acids must be recovered for the sample of interest. This is a critical challenge when working in deep subsurface environments due to the low amounts of starting biomass and the complex nature of the mineral matrix, which may establish undesired interactions with both intact cells and nucleic acids.

There are two main strategies for the isolation of nucleic acids from deep subsurface samples (see Ogram, 1998 for a more thorough review*)*. The first one consists in separating the cells from the mineral matrix by successive washings and differential centrifugations and then lysing the cells and isolating their nucleic acids (Torsvik, 1980). It results in cleaner extracts with lower amounts of PCR inhibitors (such as humic acids or other refractory organic matter), but it is strongly biased against microbes that are tightly bound to the mineral matrix (Holben, 1994). On the other hand, direct extraction procedures aim to lyse the cells (either enzymatically or mechanically) in the presence of the matrix (Zhou *et al.*, 1996). This is, in principle, a less biased approach, provided that an appropriate cell lysis protocol is applied. However, nucleic acids can be adsorbed to the particulates in the sample – especially in samples rich in silicates or iron and aluminium oxides – which greatly reduces the yield (Greaves *et al.*, 1969; Direito *et al.*, 2012). Additionally, iron minerals such as pyrite can produce free radicals when in aqueous suspension, resulting in oxidative damage to the nucleic acid molecules (Cohn *et al.*, 2006). Several solutions to these problems have been proposed, including the addition of chelating agents (Krsek & Wellington, 1999) or the neutralization of absorption sites with high concentrations of phosphate (Direito *et al.*, 2012). There is, however, no silver bullet methodology for the isolation of nucleic acids from subsurface samples

and, given the heterogeneous nature of such environments (see **section 1.2**), it may be necessary to adapt the existing protocols on a per-case basis in order to achieve optimal results.

Even when the right protocol is applied, it is unlikely that the extraction yield will be enough for downstream applications such as PCR or sequencing, as the amount of biomass in deep subsurface samples is usually very low. Therefore, an intermediate step of nucleic acids amplification is often required. Multiple displacement amplification (MDA), which uses a high fidelity polymerase in a single isothermal step, remains the method of choice for obtaining sizeable amounts of DNA from a minute quantity of starting template, even single DNA molecules (Braslavsky *et al.*, 2003). It has to be noted, however, that its use is not exempt of caveats. MDA is known to introduce amplification biases, resulting in the over- or under-representation of different microbial taxa in an often unpredictable fashion (Yilmaz *et al.*, 2010). While the use of MDA might be unavoidable when working with deep subsurface samples, special take must be taken when drawing conclusions from the quantitative results obtained with this technique.

1.4.6.4. Metaproteomics

Similar to what happened to DNA sequencing, the recent improvements in mass spectrometry proteomics in terms of robustness, ease of use and affordability have made it accessible to the general scientific community, allowing for the analysis of the full range of proteins present in a microbial community (Bruce *et al.*, 2013). While it can be used for *de novo* identification of peptides, it is much more common to use a previously obtained metagenomic library to generate a set of theoretical peptide spectra against which to score and correlate the experimental MS/MS spectra (Abraham *et al.*, 2014). With the use of shotgun metaproteomics, the protein abundances across two or more conditions cab be compared, thus adding a quantitative dimension to metagenomic data. This approach is however relatively recent (Wilmes *et al.*, 2008) and to the best of our knowledge it has not yet been applied to deep subsurface environments.

1.4.6.5. Antibody microarrays

Immunological techniques have been used for a long time for the detection and identification of microorganisms (Hock, 1995), including the design of antibodies against environmentally relevant enzymes (Bartosch *et al.*, 1999) or bacterial strains (Delehanty and Ligler, 2002).

Antibody microarrays can pack several hundreds of these antibodies in a small area, allowing for a quick and sensitive determination of the key metabolic and taxonomic features of a given sample (Parro, 2010). Albeit the amount of information they can provide is limited, and their actual performance depends on complex factors such as the generation of a set of comprehensive and non-

cross-reacting antibodies (Rivas *et al.*, 2011), if properly designed they can quickly and robustly determine the main microbiological and metabolic features of a sample, even in challenging environments such as the deep subsurface (Blanco *et al.*, 2014).

2. OBJECTIVES

In the last two decades, the Iberian Pyrite Belt subsurface has gained increasing recognition as a valuable analogue for different early Martian and early Earth environments, as well as for being the key responsible for the unique conditions found in the Tinto River waters. However, the current knowledge still suffers from a lack of consistent data on the biodiversity and metabolic potential of the IPB subsurface microbiota. Deep subsurface samples are especially complex to analyse, due to their inherent heterogeneity, low biomass, and the negative effect of mineral matrices on the performance of several microbial ecology procedures. Therefore, in order to characterise the biogeochemical processes operating in the deep subsurface of the IPB, we have focused on accomplishing the following objectives:

- To develop and validate suitable protocols for the analysis of rocky deep subsurface samples obtained by continental drilling.

- To characterise the microbial diversity and metabolic potential of the Iberian Pyrite Belt subsurface, by using both molecular and culture-based approaches.

- To identify the main challenges posed to microorganisms in deep subsurface ecosystems, and the metabolic characteristics required to overcome them.

- To integrate biological data with the physicochemical features of the samples, in order to build a model of the biogeochemical processes operating in the deep subsurface of the Iberian Pyrite Belt

- To find evidences of the underground bioreactor proposed as the origin of the unique Tinto River conditions, and to determine its limits.

3. CAMPAIGN OVERVIEW

3.1. Introduction

As discussed before, the IPB is an excellent environment to study deep subsurface ecosystems and their influence in surface processes (see **section 1.3.1**). Additionally, due to its unique geomineralogical characteristics it can be used as a model system to test hypotheses about potential past and present Martian subsurface habitats (**section 1.3.3**) and the iron-sulphur-driven early Earth biosphere (**section 1.3.2**). The overall heterogeneity of the continental crust subsurface (see **section 1.2.3**) translates into a higher possibility of finding distinct habitats when sampling both horizontally (different boreholes within the same area) and vertically (different depths within the same borehole), increasing the potential scientific return of a drilling campaign.

The scientific interest of the Tinto River area has been widely recognized by the scientific community, resulting in a large body of work (see Amils *et al.*, 2014 for the latest review). This includes two large deep subsurface drilling campaigns: the Mars Analog Rio Tinto Experiment (MARTE) project campaign (2003-2005), a joint project between the Centro de Astrobiología (CAB; INTA-CSIC) and NASA; and the Iberian Pyrite Belt Subsurface Life (IPBSL) project campaign (2011-2012). The two campaigns had different scientific goals (see below) and in some cases used different analysis techniques (due to advancements in the state-of-the-art from 2003 to 2012). However, both of them yielded an extensive biological, chemical and geological characterisation of the selected sampling sites.

This thesis is based on work performed from 2010 to 2015, and is focused in the characterisation and comparison of the microbial ecosystems present in two boreholes: the borehole 8 (BH8, drilled in the context of the MARTE project) and the borehole 10 (BH10, drilled in the context of the IPBSL project). Additionally, a novel bacterial species was isolated from a third borehole, the borehole 11 (BH11). At the beginning of this thesis, data acquisition for the BH8 was already completed, so the emphasis was put on data analysis and its integration into a comprehensive model of the ecosystem. In contrast, BH10 was drilled in the first quarter of 2012, and the results presented here derive from the author's field and laboratory work.

3.2. The Mars Analog Rio Tinto Experiment (MARTE) project

The MARTE project was a joint collaboration between CAB and the NASA Ames Research Center that aimed to simulate a robotic drilling mission to search for subsurface life on Mars (Stoker *et al.*, 2008). As such, the emphasis was put in technology development and protocol testing, culminating in a fully robotic characterisation of a shallow subsurface Mars analogue (Bonaccorsi & Stoker, 2008; Parro *et al.*, 2008). Several boreholes were performed near the Tinto River headwaters at Peña de Hierro (Spain), aiming for the sites in which groundwater intesected the massive sulphide minerals. Four of the boreholes (BH1, BH4, BH7 and BH8, reaching 40, 160, 10 and 160 mbsf respectively) were further analysed in order to search for evidences of subsurface life.

Fernández-Remolar *et al.* (2008) found evidences of biological alteration in the subsurface pyrite of BH4 and BH8, which led to to the modelling of a hypothetical Martian subsurface biosphere and its mechanisms to increase the habitability of a seemingly hostile environment (Fernández-Remolar *et al.*, 2008b). Their conclusions were however based only on geochemical and mineralogical data, since little biological information was available at the time of publication.

The microbial communities at various depths of BH8 were later characterised by a multi-technique approach that included DNA extraction followed by 16S rRNA gene amplification (and its subsequent analysis by either cloning and sequencing or hybridization into an oligonucleotide microarray described in Garrido *et al.*, 2008), CARD-FISH, and immunoprofiling with a microarray bearing 300 antibodies against ecologically relevant microbial strains, functional proteins and environmental extracts (described in Rivas *et al.*, 2008; Parro *et al.*, 2011).

3.3. The Iberian Pyrite Belt Subsurface Life (IPBSL) project

The MARTE campaigns successfully gathered evidences of an active microbial ecosystem in the IPB subsurface, but its focus was strongly biased towards technological development (see above). Additionally, in the last decade there have been impressive advances in nucleic acids sequencing technologies (**section 1.4.6**), which now allow for an unprecedented level of detail in the characterisation of microbial ecosystems.

The IPBSL project (CAB – Technische Universität Braunschweig) was designed to answer the questions left open by the MARTE project by using state-of-the-art techniques. Two boreholes (BH10 and BH11, reaching 613 and 340 mbsf respectively) were drilled in the Peña de Hierro area.

As of today, the geochemistry of the BH10 has been fully characterised by mineralogical (XRD), elemental (ICP-MS), pore water (spectrophotometry, ion chromatography, gas chromatography), petrographic and stable isotope analyses. The microbiota was in turn studied by using enrichment cultures, antibody microarrays (see previous section), massive sequencing (of the 16S rRNA gene and of a full deep subsurface metagenome) and CARD-FISH.

3.4. Borehole locations and geological settings

4.4.1. The Peña de Hierro area

Peña de Hierro is an abandoned mine site located in the catchment area of the Tinto River (**Figure 3.1**) that bears deposits comprised of a pyrite ore with variable concentrations of copper (between 1 and 4 wt%), and lower concentrations of Zn and Pb (Pinedo-Vara, 1963; Tornos, 2008). The mine was working between the middle of the nineteenth century and 1966, and it produced about 3 million tons of waste rock and other residues that were dumped around the pit, covering 25 hectares. As a result of the mining activity there is a 120 m-deep pit, of which the last 30 m are covered in acidic waters (Stoker *et al.*, 2008).

The Tinto River origin is located just south of Peña de Hierro, where waters with high ionic loads and low pH emerge from the underground as seeps. Unlike in typical acid drainage systems, where acid is generated by the exposure of sulphide minerals to the surface during mining activities (Jacobs & Testa, 2014), the acidity in the Tinto River is thought to be originated in an underground biological reactor located in the subsurface of Peña de Hierro (Gómez-Ortiz *et al.*, 2014).

3.4.2. Borehole 8

The BH8 borehole is 160 metres deep, and is located near the Peña de Hierro pit, at the GPS coordinates of N37º 43' 35.55", W6º 33' 11.60" and at an elevation of 510 metres above the sea level (masl). It cuts through a 40 m thick unit of gossan, below which sits a stockwork complex with massive veins of pyrite. At the time of drilling, the water table was situated at approximately 90 mbsf, and the aquifer was composed of oxidised waters in equilibrium with the meteoric waters recharging the Peña de Hierro pit (Fernando Tornos, personal communication). Iron oxides were detected in the first ten metres below the water table, followed by 50 metres of non-weathered stockwork. In the lowest part of the borehole (>150 mbsf), the stockwork complex was fractured, and the pyrite showed important signs of oxidation.

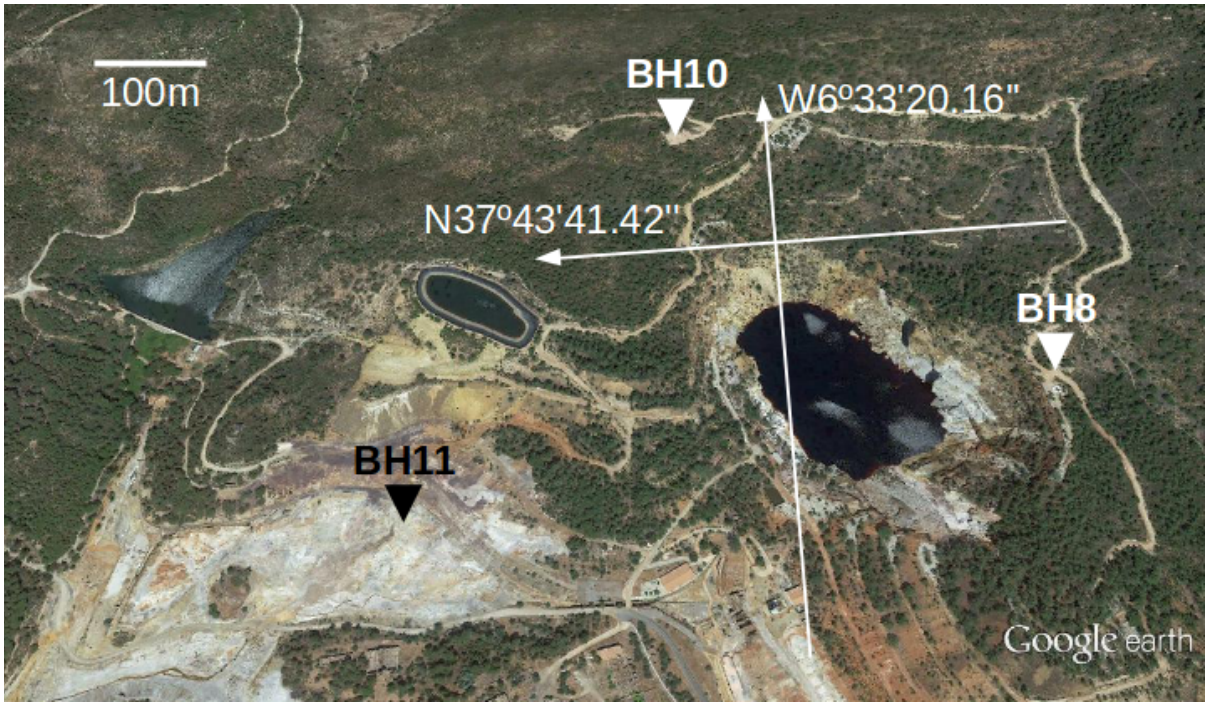


Figure 3.1. Map of the Peña de Hierro area, showing the locations of the three boreholes studied in this work.

3.4.3. Borehole 10

The BH10 borehole is 613 metres deep, and is located farther from the Peña de Hierro pit than the BH8. Its GPS coordinates are N37º 43' 45.74", W6º 33' 22.37", and its elevation is of 525 masl. It cuts the upper part of the Volcanosedimentary Complex, including four superimposed dacite to rhyodacite dome complexes that host a large zone of hydrothermal alteration with widespread replacement by quartz, sericite, chlorite and pyrite. The deepest part of the borehole cuts through the shales and greywackes of the Culm Group (see **section 1.3.1**). At the time of drilling, the water table was situated at 90 mbsf (Ricardo Amils, personal communication).

Unlike the BH8, the BH10 aquifer hosts anoxic waters with a low redox potential. Stockwork pyrite do exist, but it is unoxidised and in lower concentrations. Being farther from the Peña de Hierro pit, the rocks in the BH10 are generally less fractured, resulting in a lower permeability. Localized fracture zones do however exist at the interfaces between the different stratigraphic units.

3.4.4. Borehole 11

The BH11 borehole is 336 metres deep, and is located in a plain west of the Peña de Hierro pit. Its GPS coordinates are N37°43' 32.93", W 6°33' 32.46" and its elevation is of 428 masl. The BH11 cuts a monotonous turbiditic unit from the Culm Group, which comprises shale and sandstone with pyrite inclusions. As previously described for BH10, the BH11 rocks in the BH11 borehole are generally not fractured, and its aquifer is anoxic. At the time of drilling, the water table was situated at 5 mbsf (Ricardo Amils, personal communication).

4. MATERIALS AND METHODS

4.1. Characterisation of the BH8 subsurface ecosystem

4.1.1. Drilling and sampling

During the 2004 field campaign of the Mars Astrobiology Research and Technology Experiment (MARTE) project, a 164 m depth drill-core and sampling was performed by means of a commercial system (Fernández-Remolar *et al.*, 2008). Coring was performed with a Boart-Longyear (Salt Lake City, UT) HQ wireline system that produced 60 mm diameter cores within a plastic liner. A chemical tracer (NaBr) was incorporated into the drilling fluid to identify possible sample contamination produced during drilling. Cores were retrieved in 1 metre sections, encased in plastic liners, flushed with N_2 gas at the drilling site, sealed, and transported to a field laboratory located at the Museo Geominero in the Ríotinto village. Once, in the laboratory, cores were subsampled in an anaerobic chamber as described by Fernández-Remolar *et al.*, 2008. Briefly, powder samples were extracted every metre with a hand-drill and a sterilized bit from the centre of the fresh core face, after breaking it open with a guillotine. Especial attention for sampling was taken for those parts of the cores that showed any evidence of alteration (e.g. oxidation). All elements in contact with rock cores were cleaned and sterilized prior its use in every new sample. The obtained powder was distributed in different set of tubes as a function of the analysis to be performed. Once out of the anaerobic chamber, the samples for microbiological and molecular studies were frozen (-20ºC) or immediately processed.

4.1.2. Physicochemical characterization

The concentrations of inorganic anions such as nitrite, nitrate and sulphate, and small molecular weight organic acids such as acetate, formate and oxalate were estimated by ion chromatography as described elsewhere (Parro *et al.*, 2011). Briefly, 10 g of powdered sample were sonicated (3x1 min cycles) in 20 ml of water, and the mineral particles were removed by centrifugation. Supernatants were collected and loaded into a Metrohm 861 Advanced Compact Ion Chromatographer IC (Metrohm AG, Herisau, Switzerland) undiluted or at dilutions values of either 50% or 20%, using a *Metrosep A supp 7-250* column, with 3.6 mM sodium carbonate ($NaCO_3$) as eluent. Each sample was measured at three different dilutions to ensure that concentration ranges were within the

calibration curve for each anion (see below). The measurement error of the equipment for replicate samples was <1%. The concentration ranges for each calibration curve were the following: fluoride, bromide, propionate, formate, phosphate, tartrate, and oxalate from 0.08 (lower limit) to 2 ppm (higher limit); acetate, from 0.2 to 5 ppm; chloride, from 0.4 to 10 ppm; nitrite, from 0.2 to 5 ppm; nitrate from 0.2 to 10 ppm; and sulphate, from 8 to 200 ppm. Hydrogen, methane, iron (II) and iron (III) concentration were measured in solutions, as described in Fernández-Remolar *et al.* (2008).

4.1.3. Antibodies: production, purification, fluorescent labelling, microarray production

The antibodies used in this study were purified with protein-A and fluorescently labeled with Alexa-647 fluorochrome (Invitrogen) as described previously (Rivas *et al* 2008). The 200 antibodies described by Rivas *et al.*, 2008 were printed onto epoxy-activated microscope slides (Arrayit, CA) to form the so-called LDChip200. Printing was done in such a way that 9 identical LDChip were allocated per microscope slide, regularly spaced to match with a cassette containing the same number of hybridization chambers, the so-called MAAM (Multiple Array Analysis Module) device (Parro, 2010).

4.1.5. Sandwich microarray immunoassay (SMI) with LDChip200

Powder and shards from different core samples (0.5g) were directly analysed by SMI as described elsewhere (Parro, 2010; Parro *et al.*, 2011). Briefly, the potential molecular biomarkers or cells present in the sample were extracted into 1.5-2.0 mL of buffer TBSTRR (0.4M Tris-HCl pH 8, 0.3M NaCl, 0.1% Tween 20) by means of a hand-held ultrasonicator (DR. HIELSCHER 50W DRH-UP50H sonicator, Hielscher Ultrasonics, Berlin, Germany). The sample was then filtered through a 15 μm filter and 40 μL of the filtrate were incubated in a LDChip200 containing 200 antibodies against microbial cells, environmental extracts and biological polymers (Rivas *et al.*, 2008). After a washing step to remove the non-specifically bound material and a second incubation with a mixture of 200 fluorescent tracer antibodies, the positive antigen-antibody reactions were read by exciting the fluorochrome at 635 nm and measuring the fluorescence emission (650 nm) in a commercial scanner (GenePix). The scanned images were analysed and quantified using commercial microarray analysis software (Genepix™ Pro Software, Genomic Solutions) as described elsewhere (Parro *et al.*, 2011; Blanco *et al.* 2012).

4.1.6. DNA extraction and amplification

Environmental DNA was extracted from 2 g of powdered cores by using the commercial MoBio DNA extraction kit for soil (MoBio Labs, Solana Beach, CA, USA). The obtained DNA was used as template for simultaneous PCR amplification and fluorescent labelling by using the universal primers for prokaryotic 16S rRNA gene 8F as forward (positions 8-27 of *Escherichia coli* 16S rRNA, 5'-AGAGTTTGATCATGGCTCAG) and either1492R (positions 1492-1474, 5'-GGTTACCTTGTTACGACTT) or 16SR (positions 1057-1074, 5'-CACGAGCTGACGACA GCCG) as reverse. The thermocycler was programmed as follows: 95ºC, 5min; 10 x (95ºC 20sc, 50ºC 30sc, 68ºC 60sc); 25 x (95ºC 20sc, 48ºC 30sg, 68ºC 60sc + 5 sc per cycle); 68ºC 10 min; 4ºC.

Because the quality and yield of the obtained DNA was very low, most of the samples were subjected to a round of whole metagenome amplification by the isothermal multiple displacement amplification method (MDA) with phi29 DNA polymerase (GenomiPhi DNA Amplification Kit, GE Healthcare, formerly Amersham Bioscience). The amplified product was, in turn, used for PCR amplification or directly fluorescently labeled and analysed with the PAM oligonucleotide microarray (see below).

4.1.7. Oligonucleotide Microarray hybridization

The so-called Prokaryotic Acidophile Microarray (PAM) is an oligonucleotide microarray containing specific 16S and 23S rRNA genes probes for prokaryotic acidophiles (Garrido *et al.*, 2008). Although PAM was specially developed for the detection of prokaryotic acidophiles in the extremely acidic environment of the Tinto River (Huelva, Spain), it also contains specific oligonucleotide probes for the 16S and 23S rRNA genes from other phylogenetic groups such as Alpha, Beta, Gamma, and *Deltaproteobacteria*, Low-GC-content *Firmicutes*, high-GC-content Gram-positive bacteria, *Archaea*, *Euryarchaeota*, *Chrenarchaeota*, *Methanobacteria*, *Eukarya*, etc (see **Annex 1 – Table S1** and Garrido *et al.*, 2008 for a full list). Fluorescently labeled 16S rRNA gene amplicons were checked by agarose gel electrophoresis, then purified with Qiagen PCR purification kit columns (Qiagen, CA), and set to hybridize in HibIt hybridization buffer (Arrayit Corp.) at 50ºC for 12 h in a water bath. Then, they were washed and scanned for fluorescence as previously described (Garrido *et al.*, 2008)

4.1.8. Cloning, sequencing and phylogenetic analysis

The extracted environmental DNA was used as template for 16S rRNA gene amplification by PCR, either directly of after one round of MDA amplification. Archaeal sequences were amplified by using the archaeal-specific primer 20F (TTCCGGTTGATCCYGCCRG) paired with the universal primer U1392R (ACGGGCGGTGTGTRC), while bacterial sequences were amplified as described in the previous section. Cloning into plasmid vectors, sequencing and taxonomic assignation were performed as previously described (Parro *et al.*, 2011).

4.1.9. Fluorescent in situ hybridization (FISH) and scanning electron microscopy (SEM)

Samples of rock particles from the cores were gently washed over to remove the fixative and the pore water. Drained rock samples were immersed in 0.5% agarose to avoid the detachment of bacteria from the sample particles during incubation and washing. Hybridization and staining with DAPI (4′,6-diamidino-2-phenylindole) were carried out as previously described (Amann, 1995). Cy-3, Cy-5 and fluorescein isothiocyanate (FITC)-labeled probes for fluorescence *in-situ* hybridization (FISH) were provided by Bonsai Technology (Barcelona, Spain). Catalyzed reporter deposition (CARD) was performed using the method previously described by (Pernthaler *et al.*, 2002). Horseradish peroxidase-labelled probes and fluorochromes AlexaFluor488 and AlexaFluor534 were purchased from Molecular Probes (Invitrogen). Citifluor mounting medium (Citifluor Ltd, UK) was added to preparations to avoid fluorescence fading. An Axioskop microscope (Zeiss, Germany) equipped with the proper filter set was used to visualize the FISH results. FISH and CARD-FISH results were further visualized under a fluorescence microscopy (Axioskop II, Zeiss) and LSM510 scanning confocal microscope (Zeiss) equipped with an Ar ion laser (458–514 nm) and two He/Ne lasers (543 and 633 m). Sequence probes used were EUB338 16S GCT GCC TCC CGT AGG AGT, EUB338-II 16S GCA GCC ACC CGT AGG TGT 0–35 and EUB338-III 16S GCT GCC ACC CGT AGG TGT (Amann, 1995; Daims *et al.*, 1999) specific for *Bacteria* domain and ARCH915 16S GTG CTC CCC CGC CAA TTC CT (Stahl, 1991) specific for *Archaea* domain. Scanning electron microscopy (SEM) of cores samples was performed as previously described (Fernández-Remolar *et al.*, 2008).

4.2. Isolation and characterisation of *Tessaracoccus lapidicaptus sp. nov.*, a novel microbial species from the deep subsurface of the IPB

4.2.1. Drilling and sampling

Boreholes were continuously cored by rotary diamond-bit drilling using a Boart-Longyear HQ wireline system producing 3 meters of 60 mm diameter cores. Well water was used as a drilling fluid to lubricate the bit and return cuttings to the surface. Fluids were re-circulated. To detect potential contamination of the samples, sodium bromide (200 ppm) was added to the drilling fluid as a marker. Upon retrieval from the drilling rig, cores were divided into 60 cm length pieces, inspected for signs of alteration and stored in boxes for its permanent storage and curation in the Instituto Geológico Minero de España (IGME) lithoteque in Peñaroya. Selected cores were deposited in plastic bags, oxygen was displaced with N_2 , sealed and transported to a field laboratory located at the Museo Geominero in the Ríotinto village. Upon arrival at the field laboratory, cores were placed in an anaerobic chamber (5% H_2, 95% N_2), logged and photographed. Aseptic subsamples were obtained by splitting cores with a hydraulic core splitter and drilling out the central untouched portion with a rotary hammer with sterile bits. Bit temperature was strictly controlled (40 ºC maximum) by an infra-red thermometer.

4.2.2. Strain isolation

One gram of powder sample from a core sampled at 297 m depth, which was kept stored under anaerobic conditions, was aseptically removed in a clean N_2-filled glovebox and used to inoculate a 100 ml vial of anoxic F4 medium (0.4 g/l NaCl; 0.4 g/l NH_4Cl; 0.3 g/l $MgCl_2 \cdot 6H_2O$; 0.05 g/l $CaCl_2 \cdot 2H_2O$; 0.85 g/l $NaNO_3$; 1 g/l yeast extract; 2 g/l peptone; 1 g/l glucose; 1 g/l succinic anhydride; 7.5 g/l $NaHCO_3$; 0.5 g/l KH_2PO_4; 0.5 g/l cysteine hydrochloride; 0.5 g/l Na_2S; 1 mg/l resazurin). A negative control was performed by keeping a second vial open inside the glovebox for the duration of the inoculation procedure. The two vials were sealed with a gastight rubber septum and an aluminium cap, taken out of the glovebox and incubated at 30 ºC. Growth was followed by visual inspection and microscopic examination. After 1 week, growth was detected in the inoculated vial and 1 ml of the culture was transferred to a 100 ml vial of fresh anoxic modified F4 medium (similar to F4 medium but with no peptone and 0.1 g/l yeast extract) and incubated at 30 ºC for

another week. Pure cultures were obtained by inoculating different dilutions of this second enrichment culture onto plates of modified F4 medium supplemented with agar (20 g/l) in a clean N_2-filled glovebox and incubating them at 30 ºC in an anaerobic jar. After finding that strain IPBSL-7^T was capable of aerobic growth, cells were stored on tryptone soy broth (TSB; BD-Difco) in 20 % (v/v) glycerol at -80 ºC.

4.2.3. Morphological characterisation

Cells grown in liquid TSB for 8 h at 30 ºC and 200 r.p.m. were used to examine cell morphology and motility with a light microscope (BX40; Olympus). Cell morphology was also examined with a JEOL JEM-1010 transmission electron microscope at the Centro Nacional de Microscopía Electrónica (CNME, Madrid, Spain), after preparing the samples as described elsewhere (Ferrero *et al.*, 2013), but without adding 5 % $K_3Fe(CN)_6$ in the post-fixation step. Colony morphology and pigmentation were observed after incubation on tryptic soy agar (TSA; BD-Difco) for 5 days at 30 ºC.

4.2.4. Phylogenetic characterisation

The 16S rRNA gene was amplified by colony PCR using the universal primers 27f and 1492r (Lane, 1991), with the following PCR conditions: 95 ºC, 5 min; (95 ºC, 30 s; 48 ºC, 45 s; 72 ºC, 1.5 min) x 30 cycles; 72 ºC, 10 min. The resulting amplicon was sequenced with an ABI PRISM BigDye Terminator cycle sequencing kit (Applied Biosystems) and an automatic DNA sequencer (model 3730xl; Applied Biosystems). The resulting 16S rRNA gene sequence had a length of 1434 nt, and was aligned with the SILVA SSU Ref NR 99 reference database (Pruesse *et al.*, 2007) by using SINA (Pruesse *et al.*, 2012). The ARB software package (Ludwig *et al.*, 2004) was used to identify closely related bacteria and to reconstruct phylogenetic trees. Candidate sequences for tree reconstruction were aligned by using CLUSTAL W (Larkin *et al.*, 2007). The neighbour-joining (Saitou & Nei, 1987) with Felsenstein correction (Felsenstein, 1985) and maximum-parsimony (Fitch, 1971) algorithms included in the ARB package were used for phylogenetic inference. The robustness of the reconstructed trees was evaluated by bootstrap analysis (Felsenstein, 1985b) of 1000 resampled datasets.

4.2.5. Physiological characterisation

Growth response to temperature was tested in liquid TSB medium at 8, 15, 25, 30, 37 and 40 ºC. Growth response to pH was tested in liquid TSB medium, buffered with a 10 mM citrate buffer for pH 5 and 5.5, a 10 mM potassium phosphate buffer for pH 6, 6.5, 7, 7.5 and 8, and a 10 mM HEPES buffer for pH 8.5 and 9. Salt tolerance was tested in liquid TSB medium supplemented with 1, 2, 4 or 8 % (w/v) NaCl. Bacterial growth was determined by following OD_{600} variations of duplicate cultures shaken at 200 r.p.m. for 2 days and comparing them with uninoculated controls. The effects of temperature, pH and salt concentration on bacterial growth were also determined for *T. bendigoensis* Ben 106^T (type species of the genus *Tessaracoccus*) and *T. flavescens* SST-39^T (closest phylogenetic relative of strain IPBSL-7^T), by using the same methodology described above. The tolerance to heavy metals was determined by growing serial dilutions of the strain on TSA plates amended with different concentrations of $NiSO_4$, $ZnSO_4$, $CuSO_4$, $CdSO_4$ or $CoSO_4$. The final metal ion concentration varied from 0.5 to 20 mM. Other physiological properties (substrate utilization, enzymatic activities, tolerance to antibiotics and other compounds) of the strain were examined with commercial test systems [GPIII (Biolog), API20A, API20NE, APIZym (bioMérieux)] and antibiotic discs by the DSMZ Identification Service (Braunschweig, Germany). All commercial tests were performed according to the manufacturers' instructions on cells grown in TSB at 30 ºC. The commercial tests listed above and the cellular fatty acid analyses were also performed on *T. bendigoensis* Ben 106^T and *T. flavescens* SST-39^T cells grown on TSB at 30 ºC.

4.2.6. Chemotaxonomic characterisation

Chemotaxonomic characterisation (polar lipids, cellular fatty acids, respiratory quinones, peptidoglycan structure and DNA G+C content) was carried out by the DSMZ Identification Service. Polar lipids and respiratory quinones were extracted from 100 mg freeze-dried cell material using the two-stage method described by Tindall (1990, 1990b). Respiratory quinones were extracted using methanol/hexane, followed by phase separation into hexane. Respiratory lipoquinones were separated by TLC on silica gel (art. no. 805 023; Macherey-Nagel), using hexane/tert-butylmethylether (9 : 1, v/v) as solvent. UV absorbing bands corresponding to the different quinone classes were removed from the plate and further analysed by HPLC on a LDC Analytical (Thermo Separation Products) HPLC fitted with a reverse-phase column (2 mm x 125 mm, 3 µm, RP18; Macherey-Nagel) using methanol/heptane 9 : 1 (v/v) as the eluant. Respiratory lipoquinones were detected at 269 nm. Polar lipids were separated by two-dimensional silica gel

TLC (art. no. 818 135; Macherey-Nagel). The first direction was developed in chloroform/methanol/water (65 : 25 : 4, by vol.), and the second in chloroform/methanol/acetic acid/water (80 : 12 : 15 : 4, by vol.). Total lipid material was detected using molybdatophosphoric acid and functional groups were detected using specific spray reagents, Full details are given by Tindall (2007). Fatty acid methylesters were obtained and analysed as described elsewhere (Miller, 1982; Kuykendall *et al.*, 1988; Kämpfer & Kroppenstedt, 1996). Peptidoglycan was isolated, purified and analysed according to published protocols (Schumann, 2011). DNA G+C content analysis was carried out by the DSMZ Identification Service according to published protocols (Cashion *et al.*, 1977; Tamaoka & Komagata, 1984; Mesbah *et al.*, 1989).

4.2.7. Biomineralization studies

T. lapidicaptus was inoculated into three 100 ml bottles containing 100 ml of anoxic FE medium (wt/vol: 0.25% NaCl; 0.04% NH_4Cl; 0.003% $MgCl_2 \cdot 6H_2O$; 0.005% $CaCl_2 \cdot 2H_2O$; 0.2% $FeCl_2 \cdot 4H_20$; 0.01% yeast extract; 0.085% $NaNO_3$; 0.1% glucose; 0.1% succinic anhidryde; 0.05% KH_2PO_4; 0.025% $NaHCO_3$; 0.05% cysteine hydrochloride; 0.01% resazurin; pH=6). The cultures were incubated at 30° C and examined periodically for the presence of minerals for up to 45 days after inoculation. Sterile controls consisting of uninoculated culture medium and medium inoculated with autoclaved cells were also included. The pH of the culture medium was measured at the beginning and end of the experiment with a pH indicator strip (Merk KgaA, Darmstadt, Germany).

4.2.8. Mineral analyses

The crystals were examined by X-ray diffraction (XRD) using a PANlytical X'Pert MPD PW3011/10. A JEOL JSM 6335 scanning electron microscope (SEM), equipped with a spectroscope of dispersive energy (EDX), was used for imaging and elemental analysis of single crystals. The mineral precipitates were also analysed by transmission electro microscopy (TEM) under a JEOL JEM 2100, 200KV TEM equipped with EDX. The morphology of the cells and their association to crystal precipitates were examined with a JEOL JEM-1010 (TEM) as described in **section 4.2.3**.

4.2.9. Geochemical modelling of the FE medium

The *React* software from the Geochemist's Workbench suite (GWB, Aqueous Solutions LLC,

Champaign, IL, USA) was used to model the evolution of the FE culture medium during the experiment described in **section 4.2.7**. The inorganic ions present in medium FE were introduced in GWB as calculation constraints, and the initial pH was set at 6.. Chloride was used for charge balance, and the NH_4^+- NO_3^- redox pair was decoupled in order to allow calculations to occur in the absence of oxygen. The precipitation of FeO was disallowed, as it is not stable at low temperatures. Initial simulations indicated that the precipitation of vivianite would result in the acidification of the medium, which would be inconsistent with our pH measurement. Therefore, it was determined that the increase in pH was due to microbial activity. In order to reflect this, NH_3 was added to the simulation as a reactant in increasing amounts.

4.3. The signal-to-noise problem in deep subsurface metagenomics: wet-lab and computational solutions

4.3.1. Indirect DNA extraction

Indirect DNA extraction was performed from 10 g of powdered core sample following a protocol modified from Vilchez *et al*. (2007). Briefly, the mineral powder was first covered with 50ml of saline solution (0.9% NaCL, 100 mM EDTA, pH 5.5) and incubated overnight at 4ºC in a spinning mixer. After that, the solution was decanted at room temperature for 10 minutes and the supernatant was centrifuged at 7500 rpm for 10 minutes. The resulting pellet was subjected to phenol-chloroform extraction as described in Vilchez *et al.* (2007), removing the RNAase incubation step and applying chloroform only once (instead of twice) after the phenol-chloroform extraction. All solutions were sterilized during five minutes in a GS Gene Linker UV Chamber (Biorad Laboratories, Hercules, CA, USA).

4.3.2. Direct DNA extraction

Direct DNA extraction was performed according to Direito *et al.* (2012). Briefly, 0.5 g of powdered core sample were introduced into an Ultra-Clean Bead Tube (MoBio Laboratories, Carlsbad, CA, USA), whose original buffer had been previously removed and substituted by 1 ml of phosphate buffer (sodium phosphate 1M, 15% Ethanol). After adding 60 µl of MoBio Ultra-Clean Soil DNA solution S1, the tubes were vortexed briefly and subjected to two FastPrep (MP biomedicals, Santa Ana, CA, USA) cycles (30 seconds, power setting of 5.5) separated by one minute of ice cooling.

Following this, the tubes were incubated in a thermomixer at 80ºC for 40 minutes, while shaking at 300 rpm. The MoBio Ultra-Clean Soil DNA extraction protocol was then followed from step 10 (addition of solution S2), according to the manufacturer's instructions. All solutions were sterilized during five minutes in a GS Gene Linker UV Chamber (Biorad Laboratories, Hercules, CA, USA).

4.3.3. MDA reagent decontamination tests

Multiple displacement amplification (MDA) was performed using MagniPhi Phi29 polymerase (Genetrix, formerly X-Pol Biotech, Tres Cantos, Madrid, Spain) and random hexamers with 3' two phosphorotioate bonds, according to the manufacturer's instructions. 0.05 ng of *Leptospirillum ferroxidans* genomic DNA was mixed with the enzyme buffer and the random hexamers and aliquoted into ten 0.2 ml PCR tubes, which were laid down horizontally on a GS Gene Linker UV Chamber (Biorad Laboratories, Hercules, CA, USA) with an internal area of 763 cm^2 and subjected in pairs to five different UV intensities: No UV, 15 mJ (0.02 $mJ/cm^{2)}$, 30 mJ (0.04 mJ/cm^2), 45 mJ (0.06 mJ/cm^2) and 60 mJ (0.08 mJ/cm^2).

After the treatment, one member of each pair was re-spiked with another 0.05 ng of *Leptospirillum ferroxidans* genomic DNA in order to test if the reaction mix was still capable of DNA amplification. The mix containing the template, buffer, primers and DNA was denatured for three minutes at 95 ºC and allowed to renature for five minutes at 4 ºC. A second solution containing the dNTPs and the phi29 polymerase was prepared in parallel, aliquoted and subjected to UV sterilization as described above. The aliquots of both mixes that shared the same UV-treatment were combined and incubated for six hours at 30ºC, after which the polymerase was deactivated by a ten minute incubation at 65 ºC.

4.3.4. The Poisson binomial sequence filtering algorithm

Let us suppose we have 1 sequence of length N nucleotides (nt), each nucleotide with a potentially non-equal probability p_i of being erroneous and a probability (1-p_i) of being correct. Our target is to obtain the probability of this sequence of having j erroneous nucleotides, for j = 0, 1, 2,…, N (see example in **Figure 7.2a,b**). Statistically, our problem can be analysed as the probability distribution of the number of successes in a sequence of N independent yes/no experiments with success probabilities p_1, p_2,. . ., p_N. This is equivalent to the sum S_N of N independent Bernoulli distributed

random variables $X_1, X_2, \ldots, X_N$ such that $S_N = \sum_{i=1}^{N} X_i$, where

$$P(X_i=j) = 1-p_i \text{ for } j=0,$$
$$P(X_i=j) = p_i \quad \text{for } j=1, \qquad \text{Eq. (1)}$$
$$P(X_i=j) = 0 \quad \text{for } j>1,$$

and $P(X_i=j)$ stands for the probability of obtaining j errors in nucleotide i. The stochastic variable S_N follows a Poisson binomial distribution, from where we name the method presented here.

While the probability of obtaining a sequence with j errors in a sequence, for all values of j, can be expressed explicitly (see Eq. (**SN1.2**) and its derivation in **Annex 2 - Supplementary Note 1.1**), it becomes useless in practice for moderate values of j. We explain here an alternative algorithm inspired by Butler & Stephens (1993) that allows us to calculate the error-probability distribution $P(S_N=j)$ for all j in a simple and efficient way.

First, note that if we have two random variables Y and Z, each of them taking discrete values 0, 1, 2,..., the probability of the sum $Y+Z$ of taking value j is

$$P(Y+Z=j) = \sum_{i=0}^{j} P(Y=i)P(Z=j-i). \qquad \text{Eq. (2)}$$

The algorithm results:

1. Obtain $P(X_1=j)$ from Eq. (1). Let $U = X_1$.

2. For $i = 2, 3, \ldots, N$, the distribution is obtained by following (a-c) recursively.

   (a) Calculate $P(X_i=j)$ from Eq. (1).

   (b) Calculate $P(Y+Z=j)$ from Eq. (2), being $Y = U$ and $Z = X_i$.

   (c) Let $U = Y+Z$.

3. The exact probability for the sequence under study of having j errors, $P(S_N=j)$, is given by U when $i = N$.

4. The steps (1-3) must be repeated for $j = 0, 1, 2, \ldots, j_{max}$, where j_{max} is the lowest value of j that satisfies

$$\sum_{r=0}^{j} P(S_N=r) \geq \xi$$

and $0 < \xi < 1$ is a confidence coefficient (in our case $\xi = 0.995$). Let j_ξ be the number such that the sequence has a probability ξ of having less than j_ξ errors. It is obtained interpolating the accumulated error probability of the sequence between the values $r = j_{max}$-1 and $r = j_{max}$ to obtain its exact value in $r = j_\xi$. A linear interpolation yields

$$j_\xi = j_{max} - 1 + \frac{\xi - \sum_{r=0}^{j_{max}-1} P(S_N = r)}{P(S_N = j_{max})} .$$

5. Let j_{tol} be the maximum tolerable number of errors per sequence, that is, the maximum number of errors allowed for a correct clustering. In our calculations we have fixed $j_{tol} = 2.5$. The sequence under study is discarded if $j_\xi > j_{tol}$, and accepted as correct otherwise (**Figure 7.2b,c**). At this moment, the calculation for this particular sequence is finished, and it is time to repeat the whole algorithm for the rest of the sequences of the population.

Finally, as our problem corresponds to the sum of N binomial distributions of probabilities p_i and number of trials n=1, it can be approximated to a Poisson distribution as far as N is high and p_i <<1. The Poisson approximated probability for the sequence under study of having j errors, $P(S_N=j)$, becomes

$$P(S_N = j) = \frac{\lambda^j \exp(-\lambda)}{j!} \text{ , where } \lambda = \sum_{i=1}^{N} p_i .$$

A more detailed explanation of the Poisson binomial filtering algorithm presented above and its Poisson approximation can be found in **Annex 2 - Supplementary Note 1**.

4.3.5. Algorithm implementation

Both C and Python implementations of the Poisson binomial filtering algorithm are available in GitHub (http://github.com/fpusan/moira).

4.3.6. The moira filtering pipeline

The script *moira.py* contains an implementation of the Poisson binomial filtering algorithm and performs the following tasks:

- If required, it assembles contigs from paired reads (*--paired*). The assembler is an implementation

of mothur make.contigs command (http://www.mothur.org/wiki/Make.contigs), and includes a modified version of the Needleman-Wunsch global aligner and a consensus sequence constructor. Our implementation also returns consensus quality scores, which are simply the highest quality scores for each position of the alignment.

- It collapses identical sequences and chooses the one with the highest quality as the group representative for filtering (*--collapse*). We assumed that, in spite of differences in quality, identical sequences should have the same origin, as it is unlikely that two biologically unrelated sequences become identical due to sequencing errors. Thus, if one of them has good quality, the rest should be considered as true biological sequences and be allowed into the final dataset. We have demonstrated that collapsing sequences prior to quality filtering actually helps to mitigate an important source of taxonomic bias during sequence processing (**Annex 2 - Supplementary Note 2**).

- It truncates sequences to a fixed length (*--truncate*), discarding the sequences that are smaller than the cut-off.

- It calculates the number of errors of each remaining sequence, with a given confidence coefficient (*--alpha*) and discards the ones that have more errors per nucleotide than the specified cut-off (*--uncert*). The *alpha* confidence coefficient is defined as 1 - ξ, and represents the probability of underestimating the errors present on a given sequence.

The *moira.py* script can be downloaded from GitHub (http://github.com/fpusan/moira).

4.3.7. 16S Mock Community data

Two synthetic mock microbial communities designed by the Human Microbiome Project (23, http://www.hmpdacc.org/HMMC) were used for evaluating the different filtering methods. Genetic DNA from 22 different organisms (20 bacterial, 1 archaeal and 1 eukaryotic) was mixed in known amounts, based on qPCR of the small subunit (SSU) rRNA gene, in order to generate two different mixtures: an Even mock community, in which there is a similar amount of SSU rRNA copies for each organism, and a Staggered mock community, in which the amounts of SSU rRNA of each organism are different.

The data used in this study come from publicly available libraries generated by sequencing the Even and Staggered mock communities with the Roche 454 GS FLX Titanium, the Illumina MiSeq and the IonTorrent PGM platforms. References for all the datasets used in this study are given in **Annex 2 - Supplementary Note 3**.

4.3.8. Validation of Poisson binomial filtering on Mock Community data

The script *moira.py* was used to predict the number of errors present on each sequence for all the six Roche 454 GS FLX Titanium, the four Illumina MiSeq and the two IonTorrent PGM mock community datasets. For the MiSeq datasets, contigs were first assembled from paired-end reads by applying the *--paired* flag. The *--alpha* parameter, which indicates the probability of a read having more errors than reported, was left as its default value of 0.005. Identical reads were collapsed and the sequence with the smallest number of errors was chosen as the group representative, as described above. These predicted values were compared to the true number of errors of each sequence, which was obtained by using the mothur command *seq.error*. Briefly, the sequences were aligned to a reference database made up from the true biological sequences present in the mock community (which can be found in http://www.mothur.org/wiki/454_SOP). Sequences with less than 80% alignment coverage were discarded at this step. The resulting alignment was then used to determine the true number of errors present on each sequence, as well as whether that sequence was chimeric or not. Non-chimeric sequences that nevertheless showed less than 95% similarity to their best hit in the mock reference database were aligned again against mothur's SILVA 16S reference alignment (version 98). In case said sequence showed a pairwise identity and an alignment coverage equal or greater to 95% to any sequence in the 16S reference alignment, it was considered to be a contaminant.

4.3.9. Quality filtering of 16S reads

4.3.9.1. USEARCH

Trimming of reads by quality values was performed by using the USEARCH *fastq_filter* command, as employed by Edgar (2013). Reads (for 454/IonTorrent data) or contigs (for paired Illumina data) were truncated at the first position with a quality score below 15 (*--fastq_trunqual* 15). After that, sequences were truncated to a length of 250 nucleotides (200 nucleotides for IonTorrent data), and sequences smaller than 250 nt (200 nt for IonTorrent data) were discarded (*--fastq_trunclen* 250/200).

We also tested a different method implemented in the USEARCH *fastq_filter* command, as suggested in the author's web page (http://drive5.com/usearch/manual/uparse_cmds.html). Briefly, reads (for 454 data and IonTorrent) or contigs (for paired MiSeq data) with more than 0.5 expected

errors (--*fastq_maxee* 0.5) were discarded. After that, sequences were truncated to a length of 250 nt (200 nt for IonTorrent data), and sequences smaller than 250 nt (200nt for IonTorrent data) were discarded (--*fastq_trunclen* 250/200). NOTE: During the writing of this thesis, more details on the USEARCH expected errors method were published byEdgar & Flyvbjerg (2015). A thorough discussion on the differences and similarities between the USEARCH expected errors filter and our own is available in the published version of this work (see **section 12**).

4.3.9.2. mothur

Denoising of 454 and IonTorrent reads was performed using the mothur command *shhh.flows*, which is an implementation of the PyroNoise algorithm, as recommended in the mothur SOP (http://www.mothur.org/wiki/454_SOP; http://www.mothur.org/wiki/Ion_Torrent_sequence_analysis_using_Mothur). After denoising, mothur command *trim.seqs* was used to truncate the denoised sequences to a length of 250 nt, and to discard sequences smaller than 250 nt (200nt in both cases for IonTorrent data).

Additionally, paired Illumina reads were assembled and filtered according to mothur's MiSeq SOP (http://www.mothur.org/wiki/MiSeq_SOP). After filtering, mothur command *trim.seqs* was used to truncate the contigs to a length of 250 nt, and to discard sequences smaller than 250 nt.

4.3.9.3. QIIME

QIIME's script *split_libraries_fastq.py* was used to filter Illumina forward reads, as recommended by the authors (*-r* 3 *-p* 0.75 *-q* 3 *-n* 0) in Bokulich *et al.*, (2013).

4.3.9.4 Poisson binomial filtering

The script *moira.py* was used to perform Poisson binomial filtering on 454/IonTorrent reads or contigs assembled from Illumina paired reads (--*paired*), as described above. Before filtering, sequences or contigs were truncated to 250 nt, and the sequences smaller than 250 nt (200nt in both cases for IonTorrent reads) were discarded (--*truncate* 250/200). Identical 454/IonTorrent reads or Illumina contigs were clustered together prior to quality control (--*collapse*) and the sequence with the highest quality was chosen as the group representative for quality control. 0.01 or less errors per nucleotide were tolerated (--*uncert* 0.01) with a 0.005 chance of error underestimation (--*alpha*

0.005).

4.3.9.4 Notes

For each method, paired Illumina reads were assembled as recommended by its authors.

Note that, for consistency, we have chosen the 250 nt cut-off recommended by Edgar (2013) as the fixed length for the rest of the filtering methods (except for QIIME, which works with unpaired Illumina reads that had a fixed length of 250 nt on their own), for the 454 and Illumina datasets. Since read length may have an effect in clustering and OTU accuracy, we believe that equalizing it results in more valid comparisons between the different filtering methods. In a similar fashion, the 200 nt cutoff proposed in http://www.brmicrobiome.org/#!16sprofilingpipeline/cuhd was applied to all the filtering methods for the IonTorrent datasets.

The full list of commands used for each method can be found in **Annex 2 - Supplementary Note 6**.

4.3.10. Common processing pipeline for the filtered reads

Regardless of the filtering method, the filtered sequences were subjected to a common pipeline based in mothur's recommended SOP that included the following steps:

- Sequence alignment to mothur's Silva Reference Alignment.

- Optimization of the alignment space by removing the sequences that failed to align correctly.

- Pre-clustering of similar sequences.

- Removal of chimeras with UCHIME.

- Taxonomic classification and removal of non-bacterial and unclassified sequences.

- Library size standardization (see below).

- Clustering of the remaining sequences using mothur's default average neighbour algorithm, with an OTU distance cut-off of 0.03.

- Accuracy classification of the resulting OTUs (see below).

For each sample, the libraries obtained after filtering the raw reads with the different methods were standardized to a similar size by random sub-sampling. Total number of retrieved OTUs and singletons, as well as accuracy assessment of the OTU representative sequences, were obtained by

averaging the results from 100 independent rounds of random library size standardizations followed by clustering of the resulting reads.

The full list of commands can be found in **Annex 2 - Supplementary Note 6**.

4.3.11. OTU accuracy assessment on mock communities

The accuracy of the obtained OTU representative sequences was evaluated by aligning them to a reference database made up from the true biological sequences present in the sample, as previously described by Edgar (2013). Sequence alignment was performed with mothur *align.seqs* command. If the pairwise identity of an OTU representative sequence to any sequence in the reference database was 100%, the OTU was classified as “Perfect”. If the pairwise identity was smaller than 100%, but greater or equal to 99%, the OTU was classified as “Good”. If the pairwise identity was smaller than 99% but greater or equal to 97%, the OTU was classified as “Noisy”. If the pairwise identity was lower than 95% the OTU representative sequence was aligned to mothur’s SILVA bacterial 16S reference alignment (version 98). If said sequence showed a pairwise identity and an alignment coverage equal or greater to 95% to any sequence in the 16S reference alignment, the OTU was classified as “Contaminant”. When none of the above conditions applied, the OTU was considered to be the result of either an undetected chimera or a mock community sequence with more than 3% errors, and was classified as “Other”.

4.3.12. OTU accuracy assessment on environmental communities

OTU representative sequences from environmental communities were aligned with mothur's SILVA bacterial 16S reference alignment. For each dataset and filtering method, the average similitude of the OTU representative sequences to their best hits in the SILVA alignment was calculated. This was taken as an indicator of the overall accuracy of the resulting OTUs, under the assumption that sequencing errors are more likely to decrease OTU similitude to known sequences than to increase it.

References for all the environmental datasets used in this study are given in **Annex 2 - Supplementary Note 3.**

4.3.13. Assessment of the taxonomic bias caused by the different filtering methods

Taxonomic bias was assessed by comparing the taxonomic composition of the sample before and after performing quality filtering. Sequences were classified by using the *classify.seqs* command implemented in mothur and mothur's RDP 16S rRNA reference database (version 9). Then, taxonomic composition was obtained by calculating the proportion of sequences that were assigned to each phylotype at the genus level with an 80% confidence cut-off (40% for the environmental communities). Finally, taxonomic bias was calculated as the Bray-Curtis dissimilarity between the filtered and unfiltered sequence communities. In the 454 and IonTorrent libraries from the environmental communities, a high proportion of sequences did not get classified at the genus level. Therefore, the taxonomic composition of those libraries was instead calculated at the class level.

4.3.14. Post-sequencing assessment of contaminants and MDA amplification biases

Environmental DNA was extracted and amplified as described in **sections 4.3.2** and **4.3.3**. The resulting amplification product was purified using a MicroSpin G-50 column (GE Healthcare, Little Chalfont, United Kingdom) and 16S amplicons were pyrosequenced (Lifesequencing S.L., Valencia, Spain).

For pyrosequencing, the V3-V5 region of the 16S rRNA gene was amplified using key-tagged eubacterial primers (Lifesequencing S.L., Valencia, Spain) based on the design of Sim *et al*. (2012). PCR reactions were performed with 20 ng of metagenomic DNA, 200 µM of each of the four deoxynucleoside triphosphates, 400 nM of each primer, 2.5 U of FastStart HiFi Polymerase, and the appropriate buffer with MgCl2 supplied by the manufacturer (Roche, Mannheim, Germany), 4% of 20 g/mL BSA (Sigma, Dorset, United Kingdom), and 0.5 M Betaine (Sigma, Dorset, United Kingdom). Thermal cycling consisted of initial denaturation at 94°C for 2 minutes followed by 35 cycles of denaturation at 94°C for 20 seconds, annealing at 50°C for 30 seconds, and extension at 72°C for 5 minutes. Amplicons were combined in a single tube in equimolar concentrations. The pooled amplicon mixture was purified twice (AMPure XP kit, Agencourt, Takeley, United Kingdom) and the cleaned pool requantified using the PicoGreen assay (Quant-iT, PicoGreen DNA assay, Invitrogen). EmPCR was performed and subsequently, unidirectional pyrosequencing was carried out on a 454 Life Sciences GS FLX+ instrument (Roche, Mannheim, Germany) following

the Roche Amplicon Lib-L protocol.

The resulting pyrosequencing reads were analysed as described by Schloss *et al.* (2011). The OTUs obtained using a clustering distance of 0.09 were classified according to their most-probable source using an in-house Python script. Briefly, OTU representative sequences were aligned to the nt database using BLAST, and for each OTU, the isolation sources of its 20 first hits were retrieved and used to manually classify the OTU into one of five pre-defined groups (Water, Skin-Gut, Epilithic-Dry-Permafrost, Soil-Rhizosphere-Contaminated and Unclassified). Finally, the vegan R package (Oksanen *et al.*, 2007) was used to cluster the five samples based on their degree of similarity at the 0.09 OTU level, using the Bray-Curtis calculator. Optionally, OTU abundances were subjected to logarithmic transformation using the *decostand* function. The vegan function *metaMDS* was in turn used to represent the samples in two dimensions using non-metric multidimensional scaling (NDMS). An in-house Python script was used to blend the OTU classification results and the NMDS sample ordination into a single figure.

4.4. Characterisation of the BH10 subsurface ecosystem

4.4.1. Drilling and sampling

Core retrieval and subsampling was performed as described in **section 4.2.1**.

4.4.2. Physicochemical characterisation of BH10 samples

Ion chromatography analyses were performed as described in **section 4.1.2**. pH was measured as described in Fernández-Remolar *et al.*, 2008. The amounts of occluded hydrogen, methane and CO_2 in rock pores were measured as follows: 10 grams of mineral shards were placed into 100 ml vials, under sterile and anoxic conditions. The vials were in turn sealed with a gastight rubber septum and an aluminium cap, and their headspace flushed with nitrogen gas. After a year of incubation at room temperature, it was assumed that the gases originally present in the mineral pores had reached equilibrium with the headspace. The concentrations of hydrogen, methane and CO_2 were then measured in a Bruker 450GC gas chromatographer (Bruker, Billerica, MA, USA) using a Hayesep 80/100 column (Valco Instruments, Houston, TX, USA). Values were reported as percentages of the maximum value found for all samples. The presence of Fe^{3+}, Fe^{2+} and NH_4^+ was assessed by using

the Reflectoquant system (Merck Millipore, Billerica, MA, USA), in accordance with the manufacturer's instructions. Results of the physicochemical characterization are summarized in **Annex 3 – Table S1.**

4.4.3. Total sugars, proteins and total organic carbon determination

Proteins, sugars and total organic carbon were determined as described in Parro *et al.* (2011). Results are summarized in **Annex 3 – Table S1**.

4.4.4. Lipid extraction and characterisation.

Mineral powder was extracted using a modified Bligh-Dyer method after Talbot *et al.* (2007). Samples were deposited in a 250 ml teflon bottle, submerged in a monophasic solution of 4:10:5 water:methanol:dichloromethane and disrupted with a sonicator wand (Branson Ultrasonics, Danbury, CT, USA) for 1 hour while maintained on ice. Following sonication, the bottles were shaken at 200 rpm for 1 hour and centrifuged at 1500 rpm for 15 min. The supernatant was removed and the extraction repeated twice. 10 ml of dichloromethane and 10 ml of water were added to the pooled supernatant to induce phase separation, and the organic phase was stored. The aqueous phase was extracted with 10 ml of dichloromethane two additional times, and the pooled organic phases were dried under N_2 and weighed. 2.5 µg of pregnane diol were added to each sample as an internal standard. The extracts were acetylated by dissolving them in 50 µl of acetic anhydride and 50 µl of pyridine and heating them at 60 ºC for an hour, after which they were dried and redissolved in 100 µl of dichloromethane. Samples were analyzed on a Trace 1310 GC coupled to an ISQ LT single quad mass spectrometer (Thermo Scientific, Waltham, MA, USA). The PTV inlet was operated in CT splitless mode and held at 300°C. Separation was achieved on a Rxi-5HT fused silica column (Restek, USA; 30 m, 0.25 mm ID, 0.25µm film) with a He flow rate of 1.5 ml/min using the following temperature program: 2 min hold at 40°C; 25°C/min to 120°C; 6°C/min to 320°C; 30 min hold at 320°C. The ISQ LT was operated in electron ionization mode with a 230°C source temperature, scanning a mass range of 43-800 Da with a 0.2 s dwell time. Peak area was normalized to the internal standard pregnane diacetate, assuming a 1:1 response factor. Finally, the calculated amount of fatty acids on each sample was normalized by sample weight.

4.4.5. DNA extraction, amplification and sequencing

Up to ten grams of powdered core samples were subjected to direct DNA extraction as described in **section 4.3.2**. DNA was also extracted from a drilling water sample and from a clean PCR tube, in order to account for sampling and laboratory contaminations. The resulting extract was then amplified as described in **section 4.3.3**. DNA was finally eluted in 50 µl of 10 mM Tris-HCl buffer and quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). For the construction of microbial 16S rRNA amplicon libraries, the V5-V6 hypervariable regions were PCR-amplified using primers 807F and 1050R (Bohorquez *et al.*, 2012). The barcoding of the DNA amplicons, as well as the addition of Illumina adaptors was carried out by PCR as described previously (Camarinha-Silva *et al.*, 2014). The PCR-generated amplicon libraries were sent for 250 nt paired-end sequencing on a Illumina MiSeq platform (Illumina, San Diego, CA, USA). For the construction of metagenomic libraries, 0.8µg of amplified DNA were mixed with 1x fragmentase reaction buffer in a final volume of 18µl, vortexed thoroughly and incubated on ice for 5 minutes. The fragmentation reaction was then started by mixing the samples with 2µl of NEBNext dsDNA fragmentase (New England Biolabs Inc., Ipswich, MA, USA) and carried out for 25 minutes at 37ºC. After incubation, the fragmentation was halted by the addition of 5µl 0.5M EDTA. The ensuing DNA was purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and eluted in a final volume of 35µl prior to quantification with a Nanodrop (Thermo Scientific, Waltham, MA, USA). Metagenomic libraries were prepared with the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs Inc., Ipswich, MA, USA) using ~200ng of fragmented DNA as initial input. Size selection of 400-500bp DNA library fragments was carried out using the Agencourt AMPure XP magnetic beads (Beckman Coulter Inc., Brea, CA, USA) according to NEBNext Ultra DNA Library Prep Kit instructions. Each metagenomic DNA library was sequenced with the Illumina HiSeq 2500 platform using the TruSeq S.R. cluster kit, v3-cBot-HS (Illumina).

4.4.6. 16S community profiling

For 16S community profiling, raw 16S MiSeq paired reads were assembled and quality-filtered by using the script *moira.py* (see **section 4.3.6**) with default parameters, and then subjected to a standard clustering pipeline as recommended by the mothur authors (http://www.mothur.org/wiki/MiSeq_SOP). The OTUs obtained using clustering distances of 0.03 and 0.09 were classified according to their most-probable source using an in-house Python script. Briefly, OTU

representative sequences were aligned to the NCBI nt database using BLAST, and for each OTU, the isolation sources of its 20 first hits were retrieved and each isolation source was manually classified into one of eight pre-defined groups (Water, Skin-Gut, Epilithic-Dry-Permafrost, Soil-Rhizosphere, Sludge-Wastewater-Contaminated, AMD-Metal-Hydrothermal, Subsurface and Unclassified). The relative abundance of the eight different groups in the first 20 hits was then reported for each OTU. Please note that this classification method is slightly different to the one described in **section 4.3.14**. The *vegan* R package was used to perform hierarchical clustering, non-metric dimensional scaling and principal component analysis on the community data, as well as on the environmental parameters obtained in **sections 4.4.2** and **4.4.3**. The relationships between community composition and the environmental parameters were in turn investigated by redundancy analysis. For redundancy analysis, biological variables were log-transformed and environmental variables were both log-transformed and standardized by their standard deviate, as recommended by Kenkel (2006). The significance of the results was assessed by permutation tests.

For phylogenetic analysis, the closest 16S sequence from validly described type strains and isolates was downloaded using RDP SeqMatch tool and by blastn against the 16S ribosomal RNA sequence database of NCBI. The retrieved sequences, together with our OTU representatives, were subsequently aligned by means of the SINA web-based tool (Pruesse *et al.*, 2012). These alignments were used for phylogenetic reconstruction with MEGA6 software (Tamura *et al.*, 2013) by means of the Maximum Likelihood method based on the General Time Reversible model. A total of 100 bootstrap replications were calculated.

4.4.7. Metagenomic analysis of sample BH10-420.

Two different datasets were obtained after sequencing the metagenome of sample BH10-420. In dataset 1, no mismatches were allowed in the bar code sequence, while in dataset 2 one mismatch was allowed. Both datasets were processed independently and then combined to obtain a final assembly.

For dataset 1, paired raw metagenomic reads were quality filtered with *prinseq* (Schmieder & Edwards, 2011) with the following parameters: *-derep* 14 *-lc_method* dust *-lc_threshold* 7 *-ns_max_*p 2 *-trim_ns_left* 2 *-trim_ns_right* 2 *-trim_qual_left* 20 *-trim_qual_right* 20 *-min_qual_mean* 25. The filtered dataset was subsequently assembled with SPAdes (Bankevich *et al.*, 2012) with the following parameters: *-k* 21,33,55,77,99 *--careful --sc*. Singletons generated during quality filtering were also included in the assembly.

For dataset 2, paired raw metagenomic reads were quality filtered as described above. The filtered dataset was then subjected to digital normalization using *khmer* (Brown *et al.*, 2012). Due to low coverage, a single-step normalization was performed as recommended in the author's webpage (http://khmer.readthedocs.org/en/v1.1/guide.html). In order to achieve this, the *normalize-by-median.py* script included with khmer was run with the following parameters: *-C* 5 *-k* 20 *-N* 4 *-x* 2.5e8 *–paired*.

The filtered normalized reads from dataset 2 were assembled together with the contigs obtained from dataset 1 using SPAdes with the following command parameters: *--sc -k* 21,33,55,77,99 *–careful*. The contigs obtained from dataset 1 were provided to SPAdes with the *--trusted-contigs* parameter. Singletons generated during quality filtering were also normalized and included in the assembly. Statistics for the quality filtering, normalization and assembly processes can be found in **Annex 3 – Table S2**.

The resulting contigs were submitted to MG-RAST (Meyer *et al.*, 2008) for annotation. An in-house Python script was used to include coverage information in the sequence headers, in a format compatible with MG-RAST pipeline. Additionally, the filtered and normalized paired reads obtained from dataset 2 were directly submitted to MG-RAST without being assembled. After comparing the results obtained from annotating the contigs and the unassembled reads, the latter were deemed to be more informative, and therefore were used for the rest of the analysis. MG-RAST subsystem-based classification was used for metabolic pathway reconstruction; if this led to missing enzymes for a pathway of interest, they were searched for manually using the MG-RAST metagenome overview page.

5. CHARACTERISATION OF THE BH8 DEEP SUBSURFACE ECOSYSTEM

5.1. Introduction

The borehole 8 (BH8) was drilled in 2004 in the context of the MARTE campaign (see **section 3.2**). It reached 160 mbsf, cutting through a complex of massive stockwork pyrite. Its groundwater was in equilibrium with the waters of the near Peña de Hierro pit, and it was located near the source of the Tinto River (**Figure 5.1**; **section 3.4.2**), making this borehole an excellent system to study the influence of subsurface processes in the genesis of its acidic waters (see **section 1.3.1** for an introduction to the underground reactor hypothesis for the origin of the Tinto River). While cores were retrieved throughout the whole depth of BH8, this work focused in the saturated zone below the water table (which was located at approximately 90 mbsf), since the presence of water was deemed to be determinant for the development of active microbial communities (**section 1.1.4**).

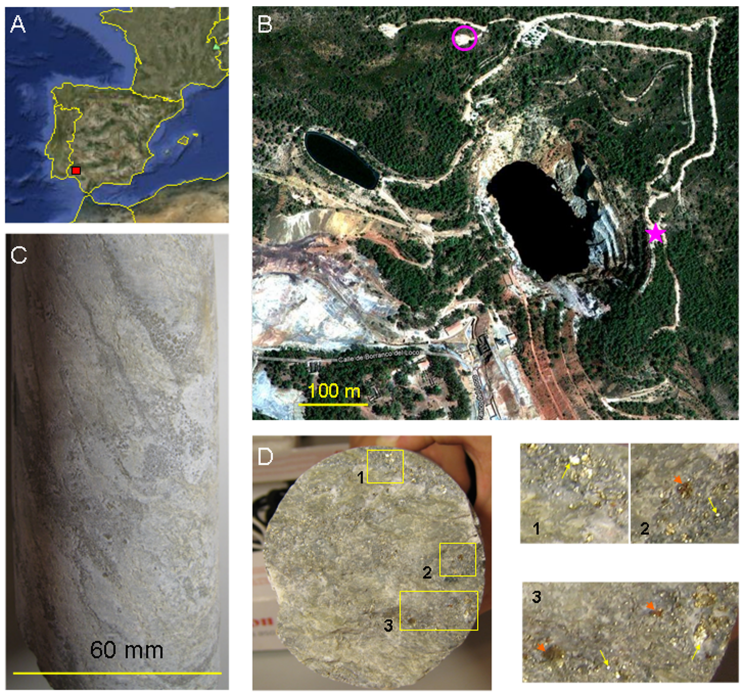


Figure 5.1. Exploration of the Iberian Pyrite Belt (IPB) subsurface geomicrobiology. Geographical location of the IPB in the Iberian Peninsula **(A)** and a satellite view **(B)** of Peña de Hierro (Nerva, Spain), where the 2004 MARTE project drilling campaign was carried out (star). The vertical distance between the drilling site and the lake water table was about 90 m. **(C)** Profile of one of the cores recovered from 130 mbs. **(D)** Cross section of the same core. Numbered squares (amplified on the right) indicate interesting features in the core, such as oxidised red spots (arrowheads) or the high abundance of bright pyrite crystals (small arrows).

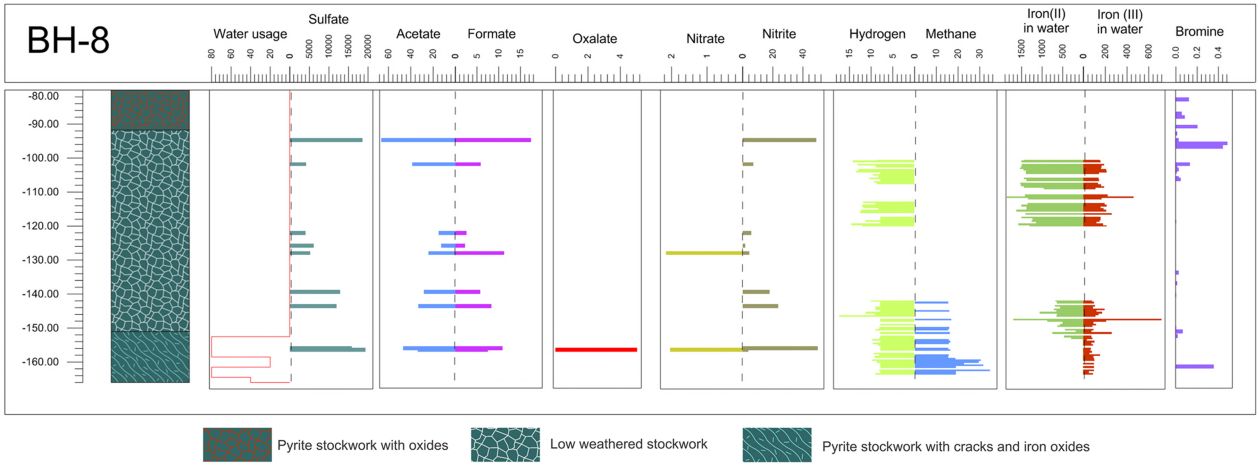


Figure 5.2. Simplified lithostratigraphical section of borehole 8 showing lithology, water usage, acetate, formate, oxalate, nitrate, nitrite and bromine concentration in rock leachates. Hydrogen, methane, iron (II) and iron (III) concentration were measured in solutions, as described in Fernández-Remolar *et al.* (2008). Sodium bromide was used as a chemical tracer included in the drilling fluid, and the bromide concentration from the rock leachate was used to calculate the contamination factor. Water usage is a measurement in the percentage of the drilling fluid loss through fractures during boring. All the concentration values are expressed in parts per million (ppm.).

5.2. Results

5.2.1. Subsurface geochemistry

Core samples were analysed by ion chromatography to identify the most relevant anions that could support microbial metabolisms (**Figure 5.2**). Nitrate and nitrite were detected along the borehole length, the latter being more abundant with peaks of > 45 µg g^{-1} at 94.7 and 156 metres below the surface floor (mbsf). High amounts of soluble sulphate were also measured, with maximal values of 18 mg g^{-1} at 94.7 and 156.4 mbsf. Especially relevant was the detection of low molecular weight organic acids, such as acetate and formate, with maximum concentrations of > 65.15 µg g^{-1} and > 15 µg g^{-1}, respectively, at 94.7 mbsf. In addition, significant amounts of oxalate were also detected in one of the samples (156.4 mbsf).

The concentration of soluble ferric iron (Fe^{3+}) was 20 to 50 µM in most of the samples, whereas that of ferrous iron (Fe^{2+}) showed a peak around 1.3 mM at 102 mbsf. The Fe^{2+} concentration was very low from 156 mbsf to the bottom of the borehole, an interval where most of the iron was in its oxidised state. Additionally, H_2 concentration was relatively constant at around 10 ppm in the anaerobic zone, while that of CH_4 increased from 142.5 mbsf to the bottom of the borehole to a maximum of 31 ppm (V).

5.2.2. Detection of microbial biomarkers using the antibody microarray LDChip200

Pulverized samples were processed by ultrasonication in an aqueous buffer and assayed with the LDChip200, an antibody microarray-based biosensor containing 200 antibodies against multiple biomolecules (Rivas *et al.*, 2008). The results indicated the presence of antigens recognized by the immobilized antibodies onto the microarray. These positive reactions corresponded to antibodies produced against antigens from Gram-positive bacteria (lipoteichoic acids, peptidoglycan), some steroid-like compounds, dsDNA, and biochemical extracts previously obtained from similar core samples (Rivas *et al.*, 2008). In addition, the LDChip200 detected ferritins and DPS-like proteins (**Figure 5.3**) in some of the samples, as well as peptides related to L-asparaginase protein.

5.2.3. Prokaryotic diversity

Total environmental DNA was extracted to determine the microbial diversity using molecular ecological techniques: (i) an oligonucleotide microarray specially designed for the detection of a broad group of prokaryotic phyla (PAM microarray; Garrido *et al.*, 2008; **section 4.1.7**); (ii) and sequencing of environmental 16S rRNA genes. The hybridization with 16S rRNA amplicons showed strong positive reactions with specific oligonucleotide probes for the Gram-positive low GC content Firmicutes phylum (LGC354a, b, c) (**Figure 5.4**; **Annex 2 – Table S2**) in some of the samples (at 120.85 and 126.85 mbsf) that previously exhibited a strong positive reaction with the anti-Gram-positive bacterial antibodies (**Figure 5.3**). Among the members of the *Firmicutes* phylum that could contribute to these signals is the genus *Desulfotomaculum* spp, which gave a clear positive in the sample at 120.85 mbsf (probe DFMI229). Other positive reactions corresponded to high GC-content Gram-positive bacteria (HGC236), the actinobacterial genus *Ferrimicrobium* (probe FRM032), some *Acidithiobacillus* spp (*A. ferrooxidans* and *A. caldus*, probes ACT465a and THC642 respectively), *Thiomonas* spp group 2 (TM2G138), *Desulfovibrio* spp (DSV698), Euryarchaeota (ANME2c760) and Methanococcales (MC1109) (**Annex 2 – Table S2)**. The hybridization of the PAM microarray with the whole MDA-amplified metagenome confirmed the previous results, and allowed the hybridization with other probes (**Figure 5.5**; **Annex 2 – Table S2**). In particular, it revealed the presence of members of the betaproteobacterial genus *Gallionella* (probe GALTS0084), the deltaproteobacterial genera *Desulfosarcina* and *Desulfococcus* (DSS658), the gammaproteobacterial WJ2-like bacterium, or the recurrent presence of high

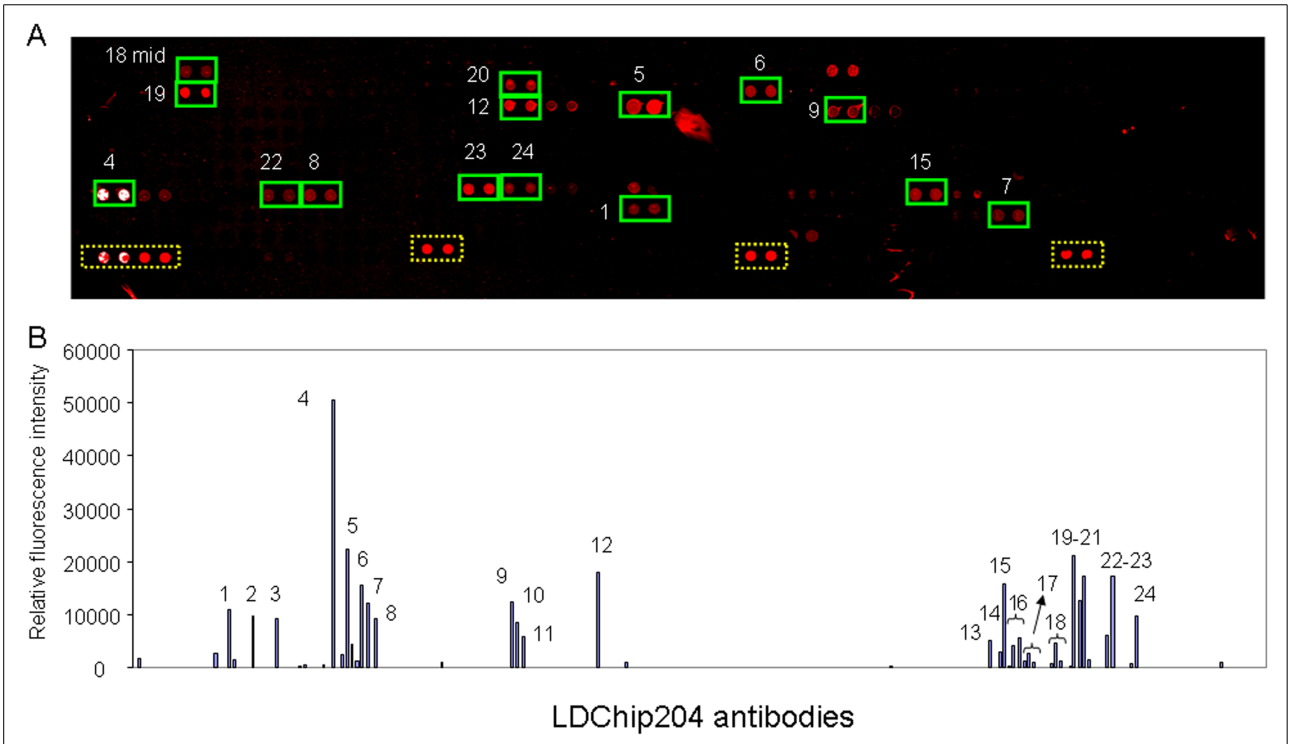


Figure 5.3. Biomarker detection in the deep IPB subsurface by the life detector chip LDChip200. A core sample (0.5 g) extracted at 150.15 mbs was analysed with LDChip 200 by sandwich microarray immunoassay as indicated in **section 4.1.5**. (A) The LDChip output is a 16 bits image with bright fluorescent spots (red to white, top picture) that showed positive antigen-antibody reactions. The fluorescent intensity can be quantified and plotted to form an immunogram (B). Due to the specificity of the antigen-antibody reaction, each positive signal in the chip immediately identify the target compound: 1, steroid-like similar to estradiol; 2, ds,ssDNA; 3, *Cyanidium* spp rich natural extract; 4, Gram-positive antigen; 5, viral antigen; 6, dsDNA; 7, ferritin protein; 8, *Hydrogenobacter* spp rich natural extract; 9, 10, 11, 12, antigens derived from extracts of IPB subsurface cores at 118-126 m, 138-139 m, 147-141 m, and 84-97 m deep, respectively (antibody names ID10S2, ID11S2, ID12S2, and ID7S2); 13, L-asparaginase protein; 14, GroEL protein; 15, bacterial Lipid A; 16, lipoteichoic acid (LTA) and *Listeria monocytogenes* LTA; 17, Lipid A, KDO bacterial antigen, and McrB protein; 18, NirS, NOR, and NRA; 19, penicillin derivatives; 20, peptidoglycan; 21, a pre-immune antiserum; 22, *Salmonella* spp; 23, cAMP; 24, tryptophan.

GC content Gram-positive bacteria (probe HGC1901, 23S rRNA gene). Among the archaea, members of the *Chrenarchaeota* and *Euryarchaeota* divisions were found (MBGB380 and ANME2c760, respectively), as well as others belonging to the *Methanosarcinae* family (MSSH859) and the *Methanococcales* order (MC1109). Finally, PCR amplified 16S rRNA gene from the samples which gave positive signals in the antibody microarray and/or PAM microarray was cloned and sequenced (**Figure 5.5**, **Annex 1 – Table S3**). Most of the retrieved sequences corresponded to the *Betaproteobacteria* (*Acidovorax*, *Comamonas*, and *Dechloromonas* genus), *Gammaproteobacteria* (*Pseudomonadaceae* family), and *Acidobacteria* divisions, which dominated

the upper (90-120 mbsf) and the lower (140-164 mbsf) part of the anaerobic zone. In the intermediate, low sulphate region of this borehole (120-139 mbsf), the sequences corresponded to the *Firmicutes* and *Actinobacteria* phyla. Additionally, several sequences associated with halophilic euryarchaea were found, in agreement with the positive hybridization with *Euryarchaea* specific probes on PAM microarray.

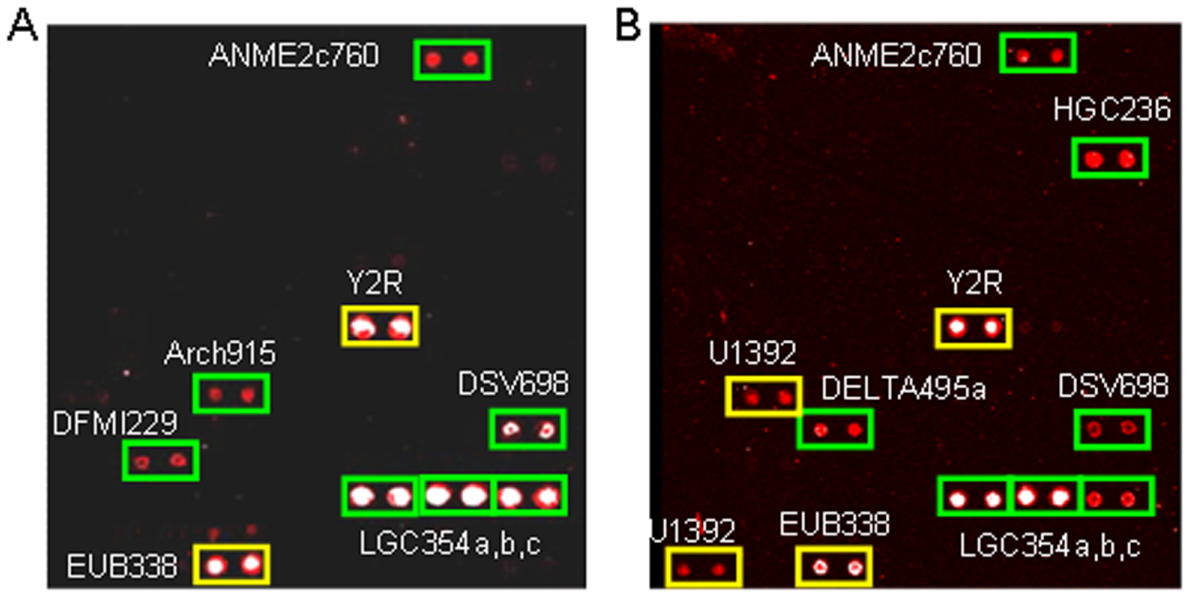


Figure 5.4. Exploring the prokaryotic diversity of the IPB with an oligonucleotide microarray. Environmental DNA was extracted from different core samples, the 16S rRNA gene was amplified by PCR, fluorescently labeled and hybridized with PAM microarray (Materials and Methods). Scanned fluorescent images obtained after hybridization of the amplified DNA from core samples 120.85 **(A)** and 126.85 mbs **(B)**. Probes showing positive results (red to white spots) are indicated: EUB338 and Y2R, universal for Eubacteria; DFMI229, genus *Desulfotomaculum* (*Firmicutes*, *Clostridiales*); DSV698, sulphate reducer *Deltaproteobacteria*, *Desulfovibrionaceae*, most *Desulfovibrio* species and *Desulfomonas*; LGC354a, b, c, Low GC Gram-positive *Firmicutes Lactobacillus* group (a), *Bacillus* group (b), and *Streptococcus* group (c); Arch915, Archaea; ANME2c760, Methanotrophic, Euryarchaeota; DELTA495a, *Deltaproteobacteria* and *Gemmatimonadetes*; HGC236, High GC content Gram-positive *Actinobacteria*.

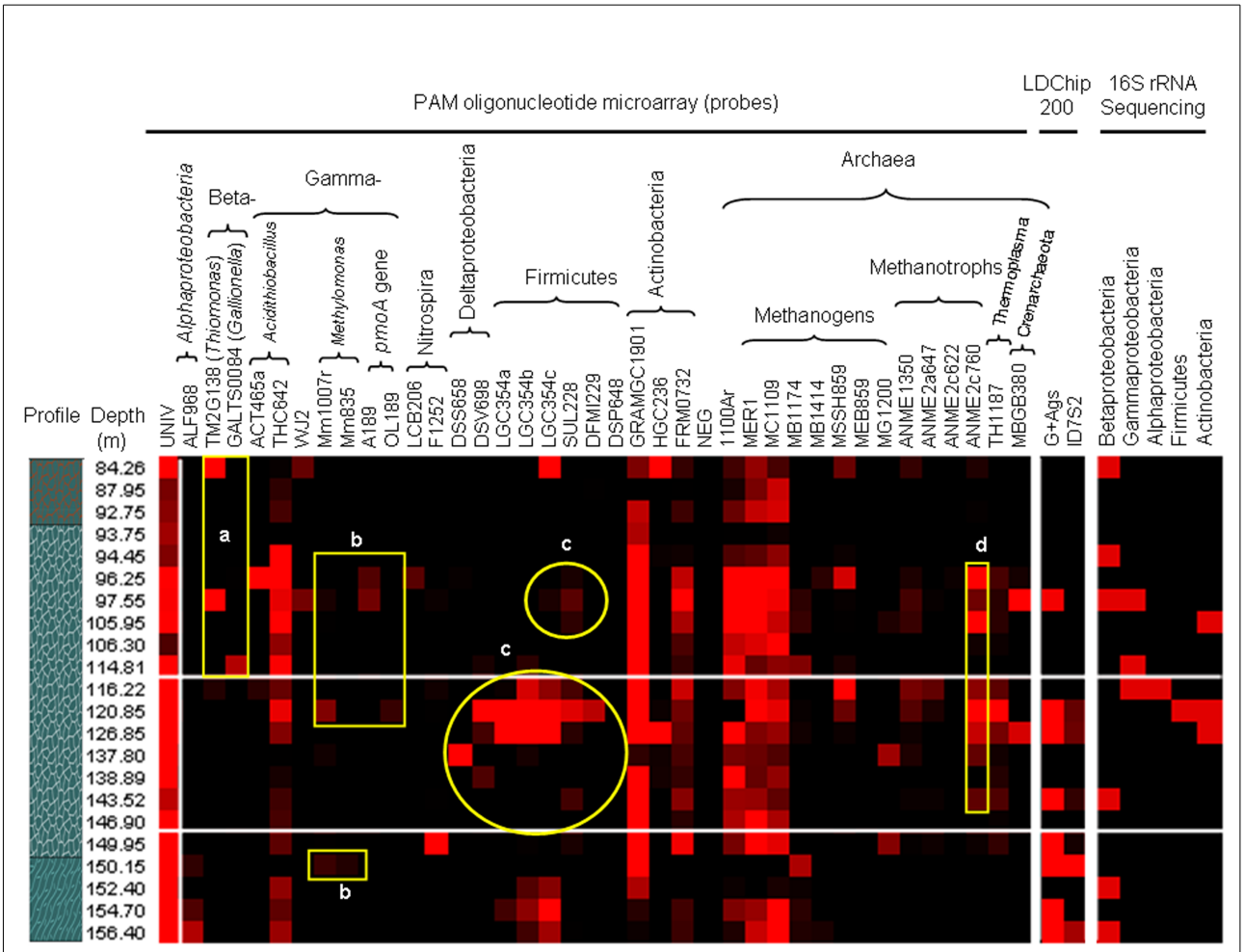


Figure 5.5. Comparison between the results obtained with three different molecular techniques: Map of relative fluorescent intensities obtained after hybridization of the PCR-amplified 16S rRNA gene and total MDA-amplified metagenome with PAM oligonucleotide microarray **(Annex 1 – Table S2)**; the fluorescent intensity of the most relevant positive biomarkers detected with the immunosensor LDChip200; and a positive (red) or negative detection (black) map of the retrieved 16S rRNA gene sequences. Each horizontal row of red and black squares represents the result for each probe obtained for a sample in the same chip. The geological profile and the depths at which the core samples were collected for analysis are indicated (left). The specificities of the PAM probes are detailed in **Annex 1 – Table S1**. The LDChip200 probes listed in this figure are the following: GP_Ag, Gram-positive bacteria antigens (mostly lipoteichoic acids); ID7S2, immunoreactive EPS material from some of the cores. Based on the prokaryotic diversity results, three geomicrobiological levels can be distinguished (white horizontal lines), which are further discussed in the text. Letters from "**a**" to "**h**" indicate clusters of probes with differentiated patterns whose distribution is further discussed in the text.

5.2.4. Metabolically active bacterial clusters

To demonstrate the actual presence of microbial cells in the IPB subsurface microniches, several core samples from BH8 were stained with DAPI and hybridized with EUB338, an eubacterial specific probe to 16S rRNA. DAPI-stained samples revealed the presence of clusters of microbial morphologies bound to micro-holes in the rock in several deep core samples (**Figure 5.6a**). Some of these clusters were also efficiently stained by FISH and CARD-FISH, indicating that the bacterial cells were indeed alive and metabolically active (**Figure 5.6b**). In addition, scanning electron microscopy (SEM) analysis of some core samples showed the presence of carbonaceous networks attached to pyrite minerals that can be originated from microbial exopolymers (**Figure 5.6c,d**) Unfortunately, while microbial clusters were indeed detected by microscopy in this study, their number was too low and their distribution too sparse to allow for a reliable quantification of the biomass in the IPB subsurface.

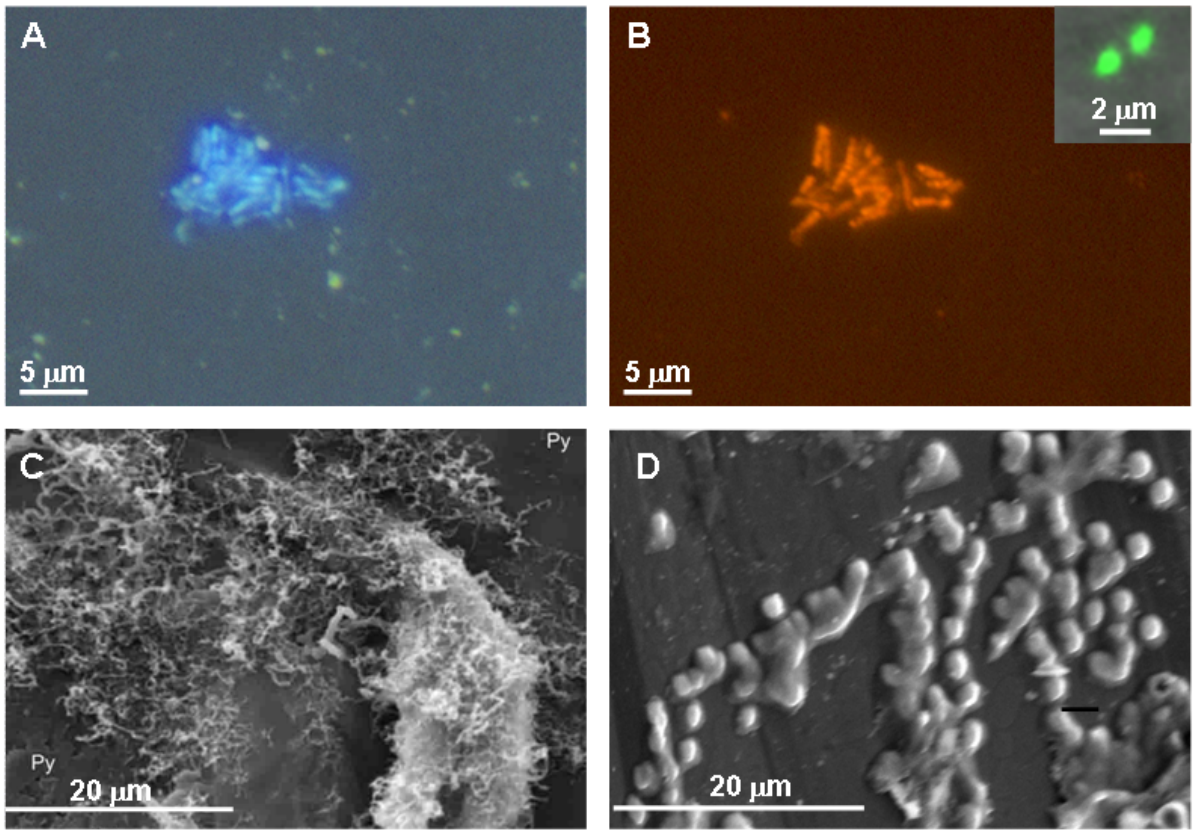


Figure 5.4. Optical (A,B) and electronic (C, D) microscopy analysis of cores from the Iberian Pyrite Belt subsurface. (**A**), (**B**) Epifluorescence micrographs showing a n eubacterial cell cluster at 154.7 mbs stained with DAPI (**A**), and the same field stained by FISH with the *Eubacteria*-specific probe EUB338 (**B**), and by CARD-FISH (insert in **B**) with the same probe. (**C, D**) Organic substances attached to the surface of pyrite minerals.

5.3. Discussion

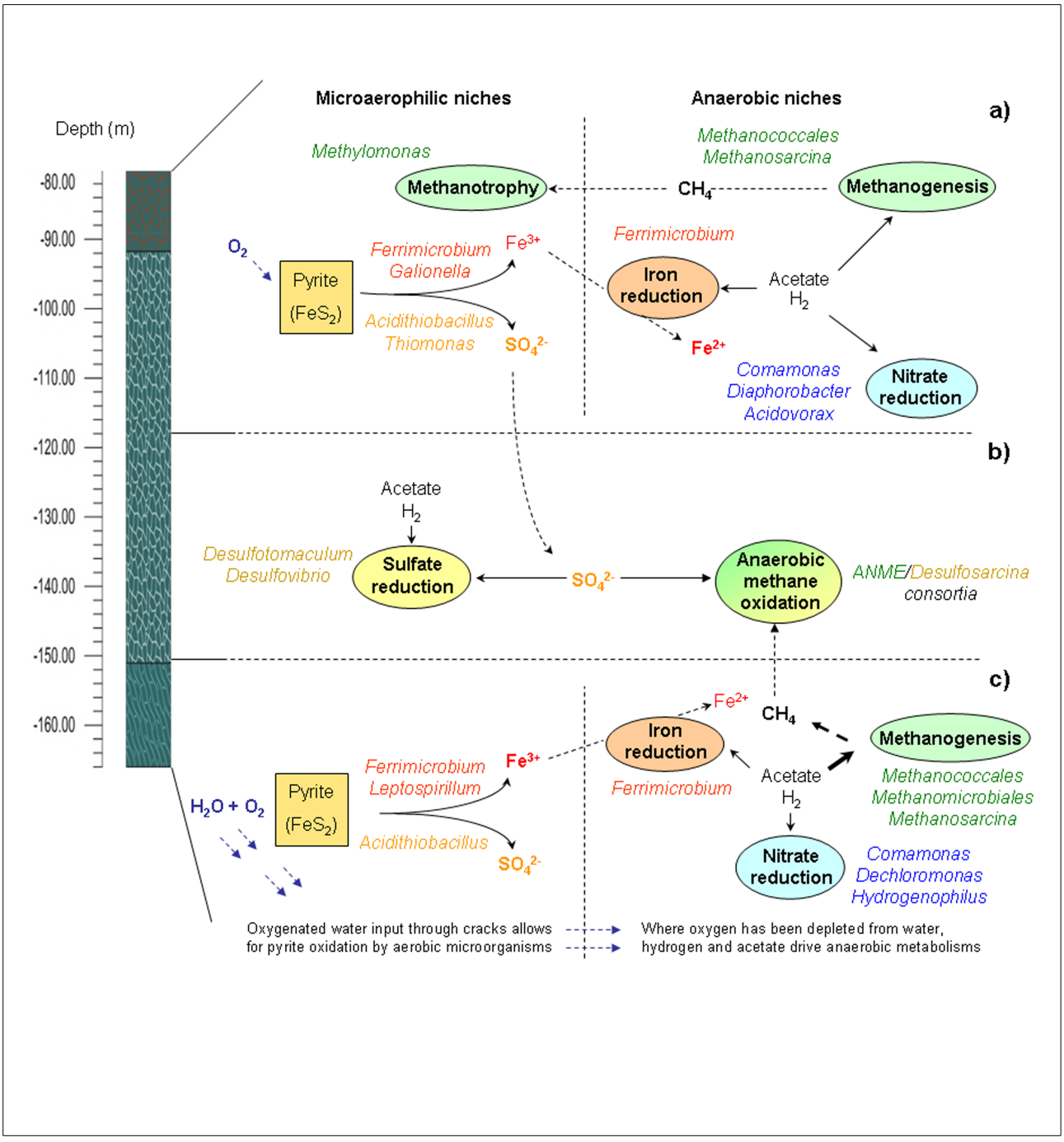
5.3.1. A multi-technique approach to explore the geomicrobiology of a deep pyrite body

Although little is known about the small-scale distribution of microorganisms in solid samples, some authors have stressed the difficulty of uniformly characterizing microbial communities in deep subsurface rocky samples (Haldeman *et al.*, 1993). Others have shown that in lacustrine sediments, microbial activities such as sulphate reduction are carried out in submilimetre microniches (Widerlund & Davidson, 2007). Such heterogeneity complicates the microbiological study of these types of samples. Also, the low concentration of microbes in the deep subsurface forces the sampling of a relatively high amount of material, especially in deep rocky samples (as in this work; **Figure 5.1**), where the microbial distribution is dispersed throughout the mineral matrix (Sahl *et al.*, 2008; Von Der Weid *et al.*, 2008). To overcome such a drawback, as well as the inherent bias of each analytical method, we retrieved many small (1-2 g) powder samples at short distances and analysed them with multiple non-culture-dependent techniques.

First, a biomarker and life detection experiment was performed by using antibody microarray-based biosensor technology. The advantages of LDChip are that: (i) the assay can be done in the field, just after sampling, which reduces the chances of sample alteration and/or contamination; (ii) very little sample processing is required, just homogenization by a hand-held sonicator, filtering and incubation; iii) up to hundreds of targets (microbes, proteins and other biomolecules) can be assayed simultaneously. Its most critical limitations concern to the cross-reaction events in a multiplex assay with a multianalyte-containing sample. We have addressed this issue by developing antibody graphs and a deconvolution method to distinguish the specific recognition events between antibodies and their cognate antigens from the cross-reaction with other related antigens (Rivas *et al.*, 2011). We concluded that although cross-reactions take place, they correspond to similar antigens or they come from phylogenetically related microbes. Another critical aspect of multiplex sandwich immunoassays is the inherent background due to the high concentration of the revealing antibody mixture, in this case 200 antibodies. Even though, we demonstrated that the system still keeps a reliable dynamic range (Rivas *et al.*, 2008). In the present work, LDChip200 mainly detected biological polymers from the Gram-positive bacteria, mostly lipoteichoic acids and peptidoglycan. One of the positive reactions corresponded to an anti-ferritin antibody, indicating the presence of bacterioferritins or the highly related DPS proteins. These proteins form part of the universal 'Ferritin-like superfamily' that plays a critical role as cellular repository of the excess of

iron (Andrews, 2010). They might be induced in the IPB subsurface as a consequence of the high iron concentration that microorganisms have to deal with. Indeed, ferritins might be playing a detoxifying role by oxidizing Fe^{2+} (highly toxic) to Fe^{3+} and storing it inside a binding pocket (Smith, 2004; Andrews, 1998). Antibodies to crude environmental extracts from cores at different depths (ID7-12S2) also showed positive reactions with a sandwich-type immunoassay, indicating the presence of complex immunogenic polymers. Whether these polymers came from Gram-positive bacteria has yet to be determined. However, the oligonucleotide microarray clearly revealed the presence of DNA from the Gram-positive *Firmicutes* phylum, and from the genus *Desulfotomaculum*, a group of obligate anaerobe sulphate reducing bacteria. Interestingly, the LDChip200 detected *Acidithiobacillus caldus* or a related strain in a sample retrieved from 156.4 mbsf (not shown) and its presence was corroborated by the positive DNA hybridization with the specific oligonucleotide probe THC642 (**Figure 5.5**). We confirmed once again the usefulness of the LDChip immunosensor for the detection of biomarkers in the subsurface, and its tremendous potential for the search for life on Mars as the key component of Signs Of Life Detector (SOLID) instrument (Parro *et al.*, 2011). The SOLID is an instrument designed and constructed for the in situ search for molecular biomarkers on Mars (Parro *et al.*, 2005; 2008; 2011). The instrument can extract the organic matter present in up to 0.5 g of dust, Martian regolith, or ground rock into a liquid solution or suspension by means of ultrasonication. After a filtering process, the liquid filtrate floods a flow chamber containing the multiplex immunosensor LDChip. After incubation and washing steps, a mixture of fluorescently labeled antibodies, a laser beam to excite the fluorochrome, and a CCD (Charged Couple Device) camera are used to reveal the positive antigen-antibody reactions. The results shown here indicate that the SOLID's life detection system works well with rocky deep subsurface samples from a Martian analogue.

The LDChip200 showed a widespread signal from Gram-positive antigens, in agreement with the positive hybridization obtained with oligonucleotide probes for low and high GC content Gram-positive bacteria with the PAM microarray. The PAM contained more probes than LDChip200, whose probe composition was highly biased for detecting microorganisms from surface environments and lacked antibodies from many groups of prokaryotes. Although the use of multiple displacement amplification (MDA) for the PAM microarray hybridization resulted in lower fluorescent signal and higher background, it corroborated the results obtained with 16S rRNA gene PCRs with DNA extracts not subjected to MDA. Furthermore, it revealed the presence of additional microbial groups. This was the case of the probe to *Gallionella*, a genus from the



Betaproteobacteria division that includes iron-oxidizing chemolithotrophs operating at low oxygen concentrations. It also revealed the presence of some members of methanogenic archaea from the *Euryarchaeota* phylum, particularly from the *Methanosarcinae* family (MSSH859) and *Methanococcales* order (MC1109), and members of the deltaproteobacterial *Desulfosarcina*/*Desulfococcus* genera.

Some of the obtained 16S rRNA gene sequences were in agreement with the results from the immunological and hybridization techniques. This technique also revealed the presence of denitrifying betaproteobacteria, who where not detected which the previous techniques. The reason of such a bias in the detection of betaproteobacteria by 16S rRNA gene sequencing compared to the other two techniques may reside in the absence of antibodies in LDChip200 and DNA probes in the oligonucleotide microarray for this phylum (only a generic probe whose target is in 23S rRNA, BET42a), and/or to the natural bias produced in the PCR amplification in natural samples with very little DNA amounts. Furthermore, MDA is known to produce biases in community composition (Yilmaz *et al.*, 2010).

The detection by SEM and FISH of clusters of microbial cells bound to minerals supported the results obtained with the techniques shown above. Although it is impossible to completely exclude the possibility of contamination of the samples with drilling fluids, the absence of the bromide tracer in the analysed samples, as well as microscopy showing the presence of mineral-bound microbial clusters, indicated that the molecular results indeed corresponded to indigenous microorganisms of the IPB.

The fact that the results obtained with the different techniques did not match exactly in every single 0.5-1g of core sample (**Figure 5.5**), can be a consequence of the heterogeneity of rocky samples, as well as the intrinsic biases of each of the current molecular ecology techniques used (Polz & Cavanaugh, 1998; Yilmaz *et al.*, 2010). In spite of this limitation, the geochemical and biodiversity analyses allowed us to cluster the samples in differentiated geomicrobiological units (**Figure 5.7**, see discussion below).

In conclusion, the microbiological analysis of environmental samples containing low biomass, such as those from deep drilling projects, requires the application of several complementary molecular techniques to obtain a more precise picture of the microbial population and the operating metabolisms. The multi-technique approach we have shown here, combining *in situ* immunosensing, multiple displacement metagenome amplification, PCR, oligonucleotide microarray hybridization and sequencing, is particularly useful to complement the biases that PCR can produce in metagenomic studies, where, very often, the amount and the quality of the template is far from the ideal.

5.3.2. Potential microbial metabolisms in the deep BH8 subsurface

Deep subsurface ecosystems are thought to be energy-limited, with low cell numbers subsisting at very low metabolic rates (Hoehler & Jørgensen, 2013). In sub-seafloor sediments the readily degradable organic substrates disappear quickly with depth, leaving only recalcitrant compounds to support heterotrophic growth (Parkes *et al.*, 2000). Altogether, the geochemical and the immunological results indicated the presence of high molecular weight and complex organic matter that can originate from extant microbial cells (protein, polysaccharides) or from extinct biomass (aromatic compounds). Small organic acids such as acetate or formate could be used as efficient carbon and energy sources, while rock-water interactions in deep bedrock fracture environments could provide the hydrogen necessary to support lithoautotrophic metabolisms (Stevens & McKinley, 1995; Freund *et al.*, 2002).

Water availability is also a key factor in subsurface environments, as it will strongly affect microbial metabolic rates and diffusion of molecular species. While most of the core samples retrieved in this work were generally compact and non-porous, the abundance of cracks in some of the deepest cores (below 150 m; **Figure 5.2**) would result in an increase in the water influx, as well as in the space available for microbial colonization. Interestingly, a diverse prokaryotic community was detected in that zone (**Figure 5.7c**), along with alterations in the geochemical profiles, such as an increase in sulphate and methane concentrations, and a decrease of ferrous iron (**Figure 5.2**), which could be products of an increased microbial activity. The presence of iron oxides and iron-oxidizing microorganisms (*Leptospirillum* spp. and *Acidithiobacillus* spp., **Figure 5.5**) in this zone also suggests that the water entering the system carries dissolved oxygen. Such oxygen would be readily consumed by these pyrite-oxidizing microbes, rendering an anoxic fluid charged with sulphate and ferric iron. This would provide an input of oxidants to the anaerobic part of the community, which contains sulphate- and iron-reducing microorganisms (see below).

In addition, methanogens were also detected (**Figure 5.5**, probes MER1, MG1200, MC1109, MSSH859), together with an increase in methane concentration (**Figure 5.2**). The co-occurrence of high sulphate and methane levels in this zone is interesting, since sulphate is known to inhibit methanogenesis (Winfrey & Zeikus, 1977; Abram & Nedwell 1978) via substrate competition from sulphate reducing bacteria (SRB). In spite of that, there was no detection of SRB below 150 mbsf, while methanogens were found in almost every sample. The reason for the predominance of methanogenesis over sulphate reduction is unclear, but it has been described that SRB can be

inhibited by high heavy metal concentrations and the precipitation of metal sulphides (Utgikar *et al.*, 2002). Ferric iron is another inhibitor of methanogenesis (Lovley & Phillips, 1987; Van Bodegom *et al.*, 2004; Zhang *et al.*, 2009). Nevertheless, Sanz *et al.*, (2011) reported methanogenic activity in microcosms from ferric iron-rich sediments of the Tinto River, as well as in *in-situ* measurements. They concluded that methanogenesis was being carried out in reduced microniches generated by iron-respiring bacteria. Our results support this hypothesis in the sense that the distribution of methanogens and ferric iron-reducers (e.g. *Ferrimicrobium* spp.) followed a similar pattern throughout the whole borehole (**Figure 5.5**, probes FRM0732, MER1, MC1109). The fact that methane was only detected below 140 mbsf in spite of methanogens being ubiquitous in the borehole can be explained by the presence of methane oxidisers in all but the deepest samples (**Figure 5.5**, **b** and **d**). Above 149 mbsf there seems to be a close association between anaerobic methane oxidizer (ANMEs) and SRB (**Figure 5.5**, **c** and **d**), which could mean that methane oxidation coupled to sulphate reduction (Orphan *et al.*, 2001; Parkes *et al.*, 2007; Knittel & Boetius, 2009; Pedersen, 2013) would be the main methane sink in the borehole (**Figure 5.7b**).

The presence of 16S rRNA sequences from denitrifying bacteria (**Figure 5.5**, *Betaproteobacteria*) is more difficult to link to the geochemical data, as the nitrate and nitrite distributions do not seem to follow a clear pattern. Nitrite peaks observed at several depths could be interpreted as the product of nitrate reduction, although their distribution is not entirely consistent with the detection of denitrifiers in the borehole.

Because the water table was located at 90 mbsf, we assume a smooth transition from an aerobic/microaerobic level to an anaerobic one (below 115 mbsf). Detection of aerobic and microaerophillic pyrite oxidisers like *Galionella* spp. and *Thiomonas* spp. above 115 mbsf (**Figure 5.5**, **a**) do in fact suggest that some aerobic microniches exist in that zone (**Figure 5.7a**). In this part of the drill microorganisms such as *Gallionella* spp might be oxidizing metals (Fe^{2+} and Mn^{2+}), and others such as *Thiomonas* spp and *Acidithiobacillus* spp could oxidise Fe^{2+} and reduced sulphur compounds (e.g. thiosulphate, $S_2O_3^{2-}$). Moreover, in those microaerobic niches, the metanotrophic bacteria *Methylomonas* spp. (**Figure 5.5b**) could account for the aerobic oxidation of methane while anaerobic metanotrophs (**Figure 5.5d**) would do the same in the anaerobic niches.

The integration of geochemical, immunological, DNA hybridization and sequencing results, allowed us to propose a three-compartment model for the deep IPB subsurface ecosystem (**Figure 5.7**): a) a microaerophilic zone (84-116 mbsf), in which methane, iron and sulphur (including pyrite, FeS_2) are oxidised in aerobic microniches, while denitrification and methanogenesis associated with ferric iron reduction occurs in the anaerobic ones; b) an anaerobic zone (115-149 mbsf), in which

anaerobic methane oxidation might be coupled to sulphate reduction; and c) a fractured zone (150-164 mbsf), in which an increased water and oxygen availability would allow for higher metabolic rates. Pyrite oxidation would consume dissolved oxygen, generating anaerobic microniches in which methanogenesis would be the main process. Thus, microbial activities would be responsible for the disappearance of reduced iron and the generation of sulphate and methane at those depths (**Figure 5.2**).

In conclusion, we showed evidence for a rich prokaryotic diversity operating in the deep subsurface of the BH8 area. The LDChip 200 immunosensor identified microbial polymers from Gram-positive bacteria as well as some proteins from the ferritin family at different depths. This result was confirmed with the PAM oligonucleotide microarray, which detected Gram-positive bacteria from the *Firmicutes* and *Actinobacteria* phyla. Additionally, PAM detected sulphate reducing bacteria, methanogenic archaea, anaerobic methane oxidizer, as well as bacteria involved in the iron/sulphur cycle. We found extensive geological, chemical and microbiological evidence of subsurface pyrite oxidation near the water table, but also in a deeper, fractured zone. Our results support the hypothesis that an underground bioreactor is responsible for the unique Tinto River conditions.

6. ISOLATION AND CHARACTERISATION OF *Tessaracoccus lapidicaptus sp. nov.*, A NOVEL MICROBIAL SPECIES FROM THE DEEP SUBSURFACE OF THE IPB

6.1. Introduction

During the course of the IPBSL campaign, anaerobic enrichment cultures for nitrate-reducing bacteria were performed. This section focuses in the enrichment, isolation and subsequent characterisation of a novel bacterial species from a 297 mbsf core sample obtained from the BH11 (see **section 3.4.4**).

6.1.1 The genus *Tessaracoccus*

The genus *Tessaracoccus*, belonging to the family *Propionibacteriaceae*, was first described by Maszenan *et al.*, (1999) and contained four species with validly published names at the time of writing: *Tessaracoccus bendigoensis* (Maszenan *et al.*, 1999), *Tessaracoccus flavescens* (Lee & Lee, 2008), *Tessaracoccus lubricantis* (Kämpfer *et al.*, 2009) and *Tessaracoccus oleiagri* (Cai *et al.*, 2011). All four species are Gram-positive and non- motile actinobacteria. The cell-wall peptidoglycan contains LL-diaminopimelic acid (LL -DAP). MK-9(H_4) is the major menaquinone and anteiso-$C_{15:0}$ is the predominant cellular fatty acid.

6.2. Results

6.2.1 Enrichment and isolation of strain IPBSL-7^T

One gram of powder sample from a core sampled at 297 m depth and stored under anaerobic conditions was aseptically removed in a clean N_2-filled glovebox and used to inoculate a 100 ml vial of anoxic F4 medium (0.4 g/l NaCl; 0.4 g/l NH_4Cl; 0.3 g/l $MgCl_2 \cdot 6H_2O$; 0.05 g/l $CaCl_2 \cdot 2H_2O$; 0.85 g/l $NaNO_3$; 1 g/l yeast extract; 2 g/l peptone; 1 g/l glucose; 1 g/l succinic anhidryde; 7.5

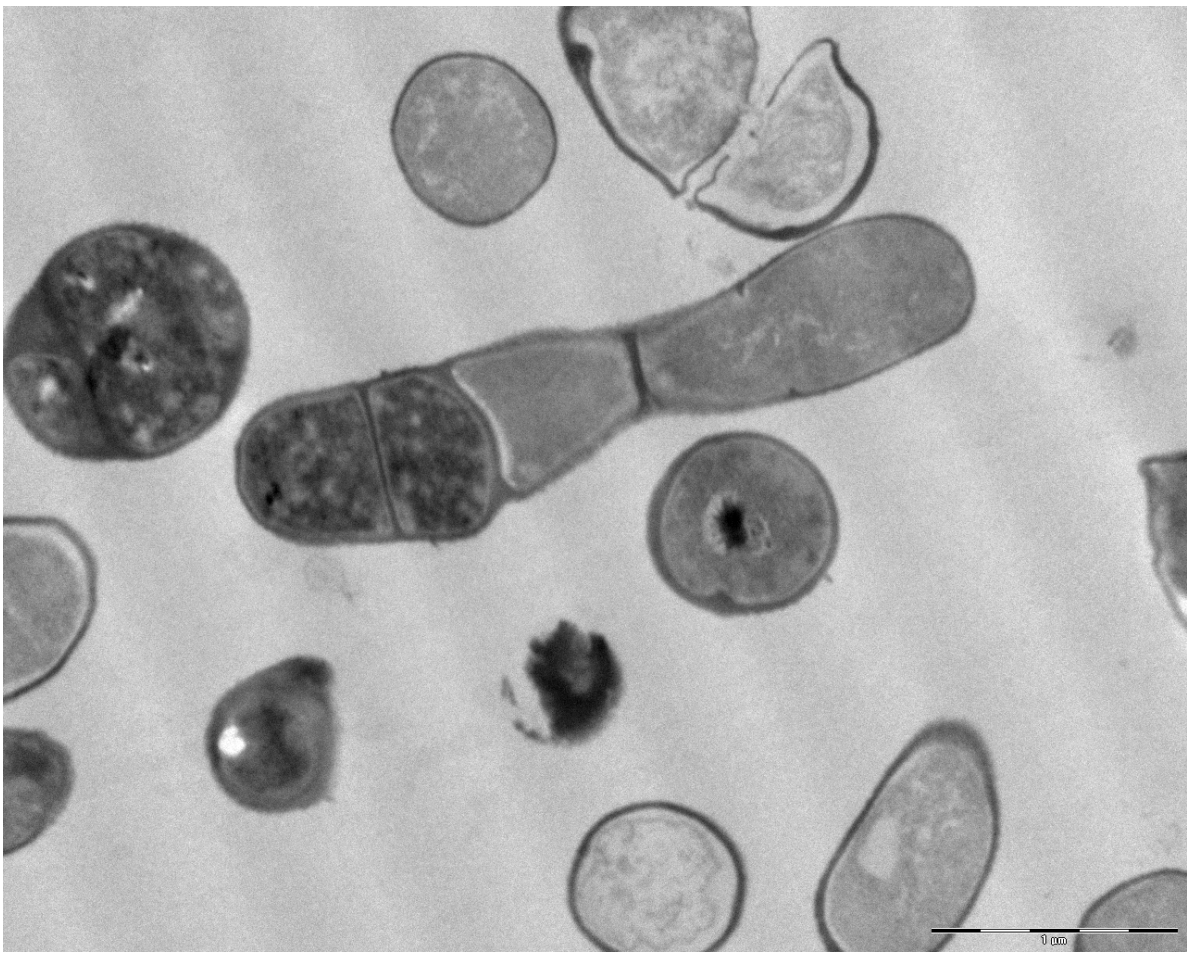


Figure 6.1. Transmission electron microscopy image of strain IPBSL-7^T cells. Bar: 1 µm.

g/l $NaHCO_3$; 0.5 g/l KH_2PO_4; 0.5 g/l cysteine hydrochloride; 0.5 g/l Na_2S; 1 mg/l resazurin). A negative control was performed by keeping a second vial open inside the glovebox for the duration of the inoculation procedure. The two vials were sealed with a gastight rubber septum and an aluminium cap, taken out of the glovebox and incubated at 30 ºC. Growth was followed by visual inspection and microscopic examination. After 1 week, growth and acetate production was detected in the inoculated vial and 1 ml of the culture was transferred to a 100 ml vial of fresh anoxic modified F4 medium (similar to F4 medium but with no peptone and 0.1 g/l yeast extract) and incubated at 30 ºC for another week. Pure cultures were obtained by inoculating different dilutions of this second enrichment culture onto plates of modified F4 medium supplemented with 20 g/l in a clean N_2-filled glovebox and incubating them at 30 ºC in an anaerobic jar. After finding that strain IPBSL-7^T was capable of aerobic growth, cells were stored on tryptone soy broth (TSB; BD-Difco) in 20 % (v/v) glycerol at -80 ºC.

6.2.2 Characteristics of strain IPBSL-7^T

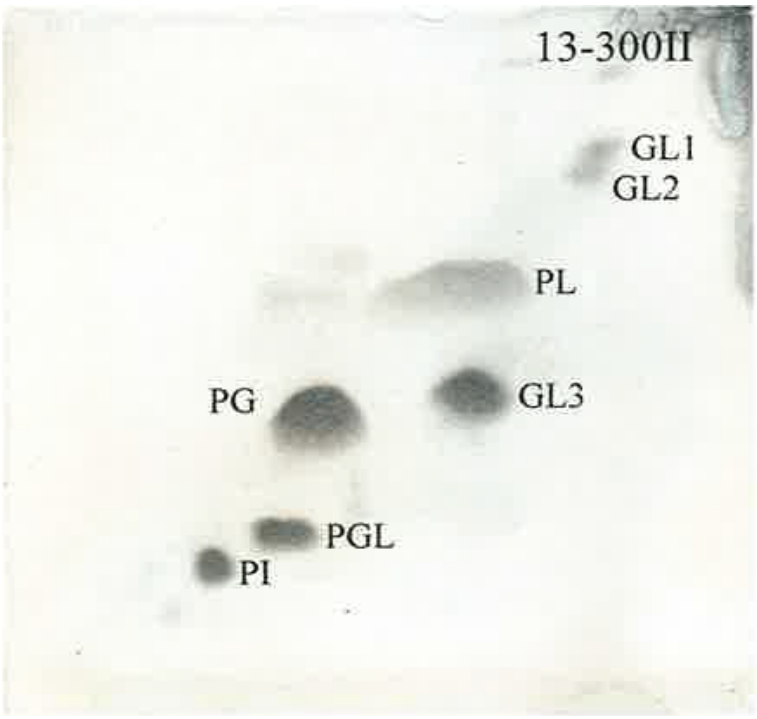


Figure 6.2. Thin layer chromatography image showing the polar lipid profile of strain IPBSL-7^T. **PGL**, unknown phosphoglycolipid; **PG**, phosphatidylglycerol; **GL**, unknown glycolipid; **PI**, phosphatidylinositol; **PL**, unknown phospholipid; **L**, unknown polar lipid.

Cells were non-motile, and pleomorphic, often occurring in pairs or small groups (**Figure 6.1**). The 16S rRNA gene sequence similarities of strain IPBSL-7^T with *T. flavescens* SST-39^T , *T. bendigoensis* Ben 106^T , *T. lubricantis* KSS-17Se^T and *T. oleiagri* SL014B-20A1^T were of 95.7, 95.7, 95.6 and 95.0 % respectively. The novel strain also showed 93.8 % 16S rRNA sequence similarity with *Brooklawnia cerclae* DSM 19609^T (Bae *et al.*, 2006), which did not belong to the *Tessaracoccus* cluster. The 16S rRNA gene sequence similarities of strain IPBSL-7^T with other type species of the family *Propionibacteriaceae* were less than 93.2 %. Results of the physiological characterisation are given in the species description and in **Tables 6.1** and **6.2**. The polar lipid profile is shown in **Figure 6.2.** Strain IPBSL-7^T was resistant to the following antibiotics (mg of antibiotic per disc, unless otherwise specified): colistin (10), nitrofurantoin (100), pipemidic acid (20), fosfomycin (50) and nystatin (100 units); weakly sensitive to aztreonam (30), gentamicin (10), amikacin (30), norfloxacin (10), polymyxin B (300 units) and kanamycin (30); and sensitive to penicillin G (10 units), oxacillin (5), ampicillin (10), ticarcillin (75), meziocillin (30), cefalotin (30), cefazolin (30), cefotaxime (30), imipenem (10), tetracycline (30), chloramphenicol (30), vancomycin (30), erythromycin (15), lincomycin (15), ofloxacin (5), bacitracin (10 units), neomycin (30), doxycyclin (30), ceftriaxone (30), clindamycin (10), moxifloxacin (5), linezolid (30), quinupristin/dalfopristin (15), teicoplanin (30) and piperacillin/tazobactam (40). Strain IPBSL-7^T tolerated up to 2 % (w/v) NaCl, 1 mM Zn^{2+}, 2 mM Ni^{2+}, 1 mM Cu^{2+} and 1 mM Co^{2+}.

Table 6.1. Differential characteristics between strain IPBSL-7^T and its most closely related phylogenetic neighbours.

Characteristic	IPBSL-7^T	*Tessaracoccus bendigoensis*^{T}	*Tessaracoccus flavescens*^{T}
Cell morphology	Oval to rods (0.45x0.5-1.0 µm)	Cocci (0.5–1.1 µm) *	Rods (0.6×1.2 µm) *
Growth temperature (range, optimum) (°C)	15-40, 37	15-40, 25	15-40, 30
Growth pH (range, optimum)	6-9, 8	5-9, 8	6–9, 8.5
NaCl concentration for growth (range, optimum) (%)	0-2, 0	0-8, 4	0–8, 0
Carbon source utilization			
Dextrin	-	-	+
D-Maltose	+	-	+
D-Trehalose	+	-	+
D-Cellobiose	+	-	+
β-Gentobiose	-	-	+
Sucrose	+	-	+
D-Turanose	+	-	+
Stachyose	+	-	+
D-Raffinose	+	+	-
α-D-Lactose	+	-	+
D-Melibiose	+	+	+
(β-Methyl-D-Glucoside)	-	-	+
D-Salicin	+	-	+
N-Acetyl-D-Glucosamine	-	-	+
N-Acetyl-β-D-Mannosamine	-	-	+
N-Acetyl-D-Galactosamine	-	-	+
N-Acetyl-Neuraminic Acid	-	-	+
D-Glucose	+	-	+
D-Mannose	+	-	+
D-Fructose	+	-	+
D-Galactose	+	-	+
L-Fucose	-	-	+
L-Rhamnose	-	-	+
Inosine	-	-	+

Characteristic	IPBSL-7^T	*Tessaracoccus bendigoensis*^{T}	*Tessaracoccus flavescens*^{T}
D-Sorbitol	+	-	+
D-Mannitol	+	-	-
D-Arabitol	+	-	+
myo-Inositol	+	-	+
Glycerol	-	-	+
Pectin	-	-	+
D-Gluconic acid	-	-	+
D-Glucuronic acid	-	-	+
Glucuronamide	-	-	+
D-Lactic acid methyl ester	-	-	+
L-Lactic acid	-	-	+
α-Hydroxy-Butyric acid	-	-	+
Acetoacetic acid	+	-	+
Acid production from			
D-Glucose	+	+	w
D-Mannitol	w	+	-
D-Lactose	-	+	-
D-Saccharose	-	+	w
D-Maltose	-	+	+
D-Xylose	-	+	+
L-Arabinose	-	+	+
Enzyme activities			
Naphtol-AS-bi-phosphohyrdolase	-	w	w
N-acetyl-β-Glucosaminidase	-	w	+
Chemotaxonomic data			
Major menaquinones	MK-9(H_4) (91%), MK-9(H_6) (4%), MK-9(H_2)(2%)	MK-9(H_4) *	MK-9(H_4) (84%), MK-8(H_0) (12%), MK-7(H_2) (4%) *
Phospholipids‡	PGL, PG, GL, PI, PL, L	DPG, PG, PI, PL *	DPG, PG *
DNA G+C content (mol%)	70.3	74 *	68.4

* Data taken from Cai *et al.*, 2011. +, Positive; −, negative; w, weakly positive. ‡ PGL, unknown phosphoglycolipid; DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; GL, unknown glycolipid; PI, phosphatidylinositol; PL, unknown phospholipid; L, unknown polar lipid.

All strains are Gram-positive, facultatively anaerobic and contain LL-DAP in the cell wall peptidoglycan. All strains are positive for nitrate reduction, esculin hydrolysis, glucose fermentation, glycerol fermentation and mannose fermentation. All strains are positive for the following enzymatic activities: esterase, esterase-lipase, leucin arylamidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase.

All strains are negative for indol production, arginin hydrolysis, urease activity and gelatine hydrolysis. All strains were unable to grow on the following substrates: 3-O-methyl-D-gGlucose, D-fucose, L-fucose, L-rhamnose, inosine, D-

sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol, D-glucose-6-phosphate, D-fructose-6-phosphate, D-aspartic acid, D-serine, gelatin, glycil-L-Proline, L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L-hystidine, L-pyroglutamic acid, L-serine, pectin, D-galacturonic acid, L-galactonic acid γ-lactone, D-gluconic acid, D-glucuronic acid, glucuronamide, mucic acid, quinic acid, D-saccharic acid, p-hydroxy-phenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, citric acid, α-keto-glutaric acid, D-malic acid, L-malic acid, bromo-succinic acid tween 40, γ-amino-n-butyric acid, α-hydroxy-Butyric acid, β-hydroxy-butyric acid, α-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, sodium formate.
All strains were negative for the following enzymatic activities: alcaline phosphatase, lipase, valin arylamidase, cystin arylamidase, trypsin, chymotrypsin, acid phosphatase, β-glucuronidase, α-mannosidase, α-fucosidase.

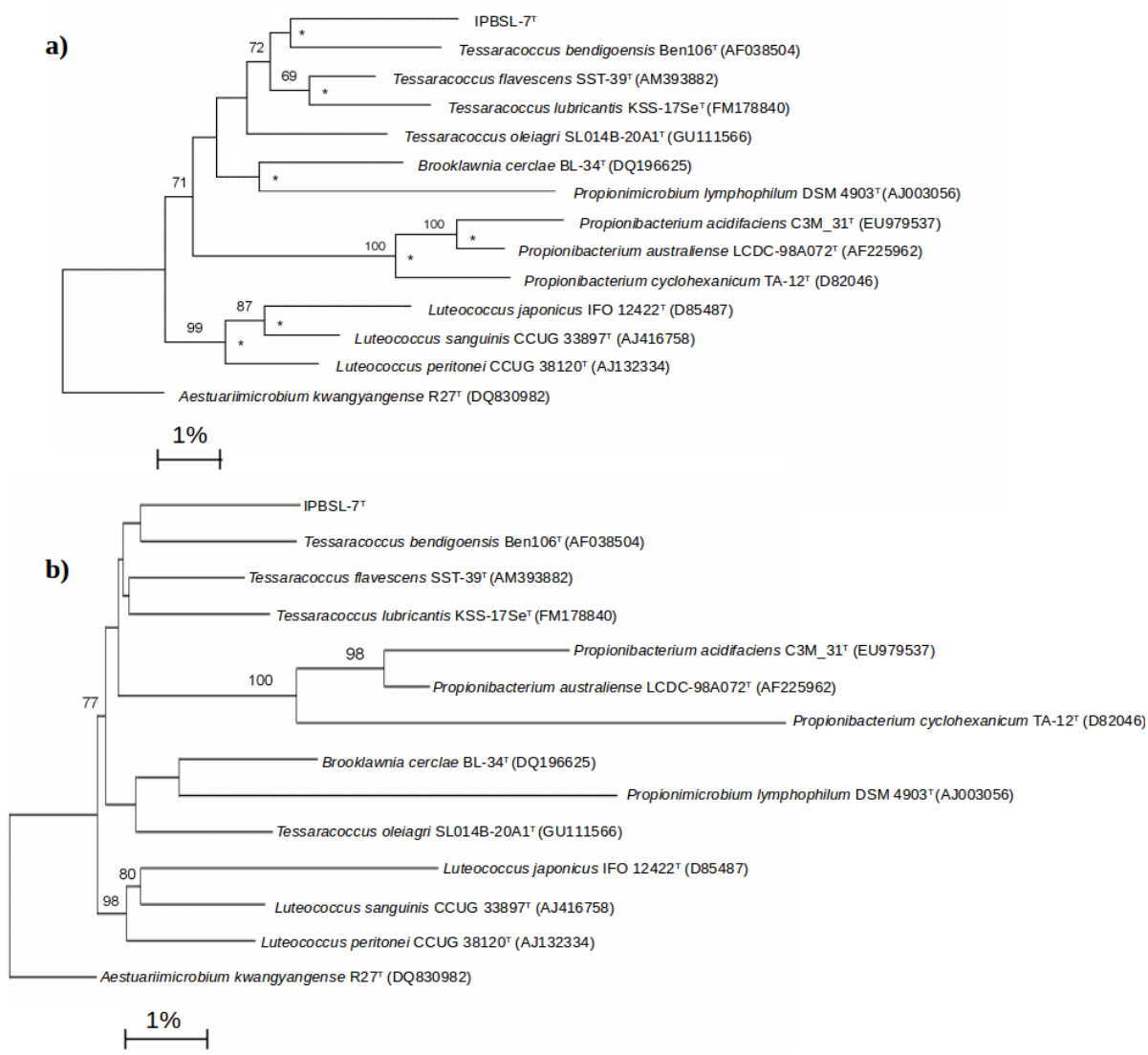


Figure 6.3. Phylogeny of strain IPBSL-7^T. a) Neighbour-joining tree based on 16S rRNA gene sequences showing the relationship of strain IPBSL-7^T with other members of the family Propionibacteraceae. Bootstrap values smaller than 65% are not shown. Nodes marked with asterisks were also found in the maximum parsimony tree. Bar: 1 nt substitutions per 100 nts. **b)** Maximum parsimony tree based on 16S rRNA gene sequences showing the relationship of strain IPBSL-7^T with other members of the family Propionibacteraceae. Bootstrap values smaller than 65% are not shown. Bar: 1 nt substitutions per 100 nts.

Table 6.2. Fatty acid composition of strain IPBSL-7^T and its most closely related strains.

Fatty acid (%)	IPBSL-7^T	*Tessaracoccus bendigoensis*^{T}	*Tessaracoccus flavescens*^{T}
Saturated fatty acids			
C12:0	0.8	0.3	-
C14:0	3.3	2.48	1.4
C14:0 2-OH	−	2	-
C15:0	-	0.6	-
C15:0 2-OH	-	0.3	-
C16:0	5	9.5	3.1
C17:0	2.7	-	-
C18:0	1	0.4	0.9
Unsaturated fatty acids			
C13:1 at 12-13	−	1.3	-
C15:1ω6c	−	-	-
C15:1ω8c	−	-	-
C16:1ω9c	1.1	-	-
C17:1ω6c	0.8	-	-
C17:1ω8c	2.6	-	-
C18:1ω7c	-	0.4	-
C18:1ω9c	1.4	1.7	-
C20:4ω6,9,12,15c	-	1.4	-
Branched fatty acids			
iso-C14:0	3.6	2.5	9.6
iso-C14:0 3-OH	−	0.8	0.7
iso-C15:0	7.5	2.5	14.2
iso-C16:0	6.4	1.2	14.7
iso-C17:0	1.9	-	1.8
iso-C18:0	0.5	-	1.1
anteiso-C13:0	−	0.5	-
anteiso-C15:0	55	48.8	46.9
anteiso-C15:1 A	0.5	-	0.6
anteiso-C17:0	2.8	0.5	1.2
anteiso-C17:1ω9c	0.8	-	-
Summed features*			
1	−	2.7	-
2	−	0.8	-
3	2.5	-	-
4	−	11.27	3
5	-	0.5	-

* Summed features are groups of fatty acids that could not be separated by GC with the MIDI system. Summed features 1, 2, 3, 4 and 5 comprised C13:0 3-OH/C15:1, C16:1ω7c/ iso-C15:0 2-OH , C16:1ω6c/ C16:1ω7c, iso-C17:1 I/anteiso- C17:1 B and C18:2ω6,9c/ anteiso-C18:0, respectively.

6.2.3. Bioprecipitation of iron phosphates and carbonates by strain IPBSL-7^T

Cells from strain IPBSL-7^T were grown anoxically in FE medium, which is enriched in iron and carbonate (**section 4.2.7**). After 45 days, the pH had changed from 6 to ~7.5 and mineral precipitates had formed in the inoculated bottles, while the sterile parallel controls (bottles with autoclaved cells and without cells) remained clear and showed no pH changes. The XRD study revealed that the bioprecipitates were composed of vivianite and siderite, with vivianite being the dominant mineral phase (**Figure 6.4**).

TEM and SEM images of the bacterial precipitates show that Fe-phosphate crystals and Fe-carbonate spheroidal nanoglobules were attached to the bacterial cells and surroundings (**Figures 6.5A-D**; **6.6A,B,D**; **6.7A,B**). Mineralized bacteria were clearly recognized (**Figures 6.6A; 6.7A**) as well as dividing or septated cells (**Figures 6.5B**; **6.7A**). EDX analyses (**Figure 6.5F-H**) confirmed the X-ray results by detecting the presence of iron phosphates and carbonates. The vivianite crystals had a prismatic or tabular habit and formed coarse radial-fibrous aggregates like rosettes with a high degree of crystallinity and vitreous lustre (**Figure 6.6C,D**). On the other hand, the siderite crystals were aggregates of nanoglobules with a diameter 20-100 nm (**Figure 6.7A-B**). These nanoglobules were attached to IPBSL-7^T cells and embedded in a thin organic film that most likely corresponded to exopolymeric substances (EPS) (**Figures 6.5A-D**; **6.7A,B**). Spherulites, or more crystalline forms of siderite, were also detected (**Figure 6.7C,D**), although they were far less abundant than the nanoglobules, which is consistent with the smaller overall crystallinity of the siderite phase (**Figure 6.4**).

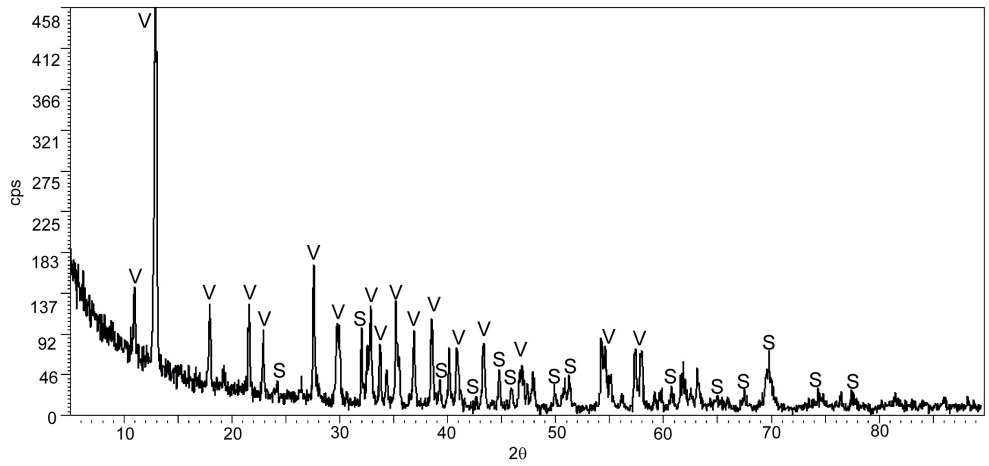


Figure 6.4. X-ray difractogram of the bioprecipiates formed in IPBSL-7^T anaerobic cultures. Vivianite (V), siderite (S).

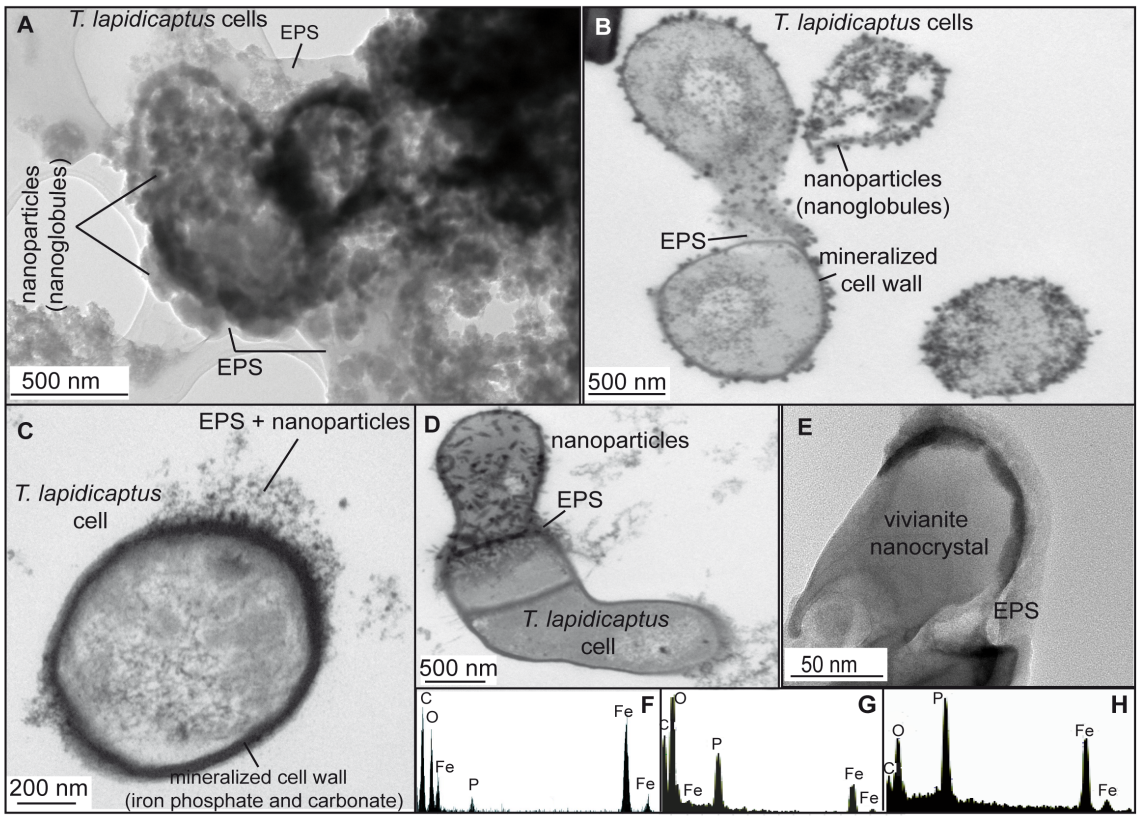


Figure 6.5. TEM images of the bioprecipitates formed in IPBSL-7^T (*Tessaracoccus lapidicaptus*) anaerobic cultures. A) Bioprecipitates attached to IPBSL-7^T cells and their secreted EPS, respectively. **B)** IPBSL-7^T cells with mineralized cell walls. **C)** Detail of IPBSL-7^T cell with a mineralized cell wall. Note the nanoparticles embedded in EPS. **D)** Detail of three cells together. The upper cell is covered by nanoparticles. **E)** Elongated nanoparticle, vivianite nanocrystal, embedded in EPS. **F,G)**: EDX spectra of both dark and lighter mineralized areas (see **A**) composed of Fe-carbonate and phosphate, respectively. **H**) EDX spectrum of a nanoparticle from **E)** composed of Fe-phosphate (vivianite).

6.2.4. Geochemical modelling of mineral precipitation in anoxic FE medium

Preliminary simulations showed that vivianite should precipitate abiotically in the FE medium, dropping the pH from six to approximately four (**Figure 7.8a,d**, 0 mmol of NH_3 added), although neither minerals nor pH changes were observed in the sterile controls. The increase in pH observed in the inoculated cultures was simulated by adding increasing concentrations of NH_3 to the model solution, after which the evolution of pH, aqueous species and precipitated minerals in the system was calculated (**Figure 7.8**). The removal of protons by NH_3 was countered by the precipitation of vivianite (and the associated decrease of phosphate in the solution), resulting in a very slow increase in pH until all the vivianite had precipitated (**Figure 7.8a,b,d**). After that, the addition of further NH_3 rapidly increased the pH to around 7.5, concurrently with the precipitation of siderite and FeO.

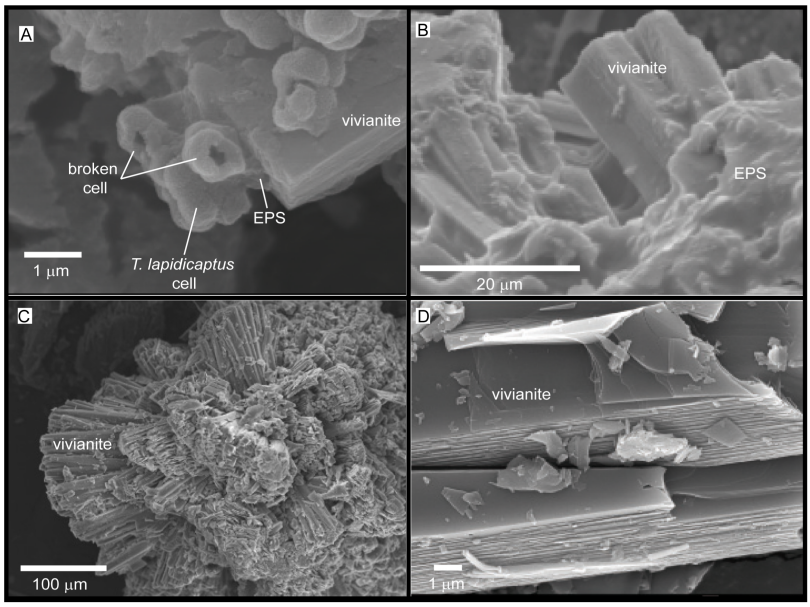


Figure 6.6. SEM images of the Fe-phosphate precipitates from anaerobic cultures of the deep subsurface strain IPBSL-7^T (*Tessaracoccus lapidicaptus)* .
A) Vivianite crystal attached to mineralized IPBSL-7^T cells.
B) Elongated vivianite crystal embedded in EPS. **C)** Rosette formation of crystal clusters of vivianite.
D) Vivianite crystals with tabular habit.

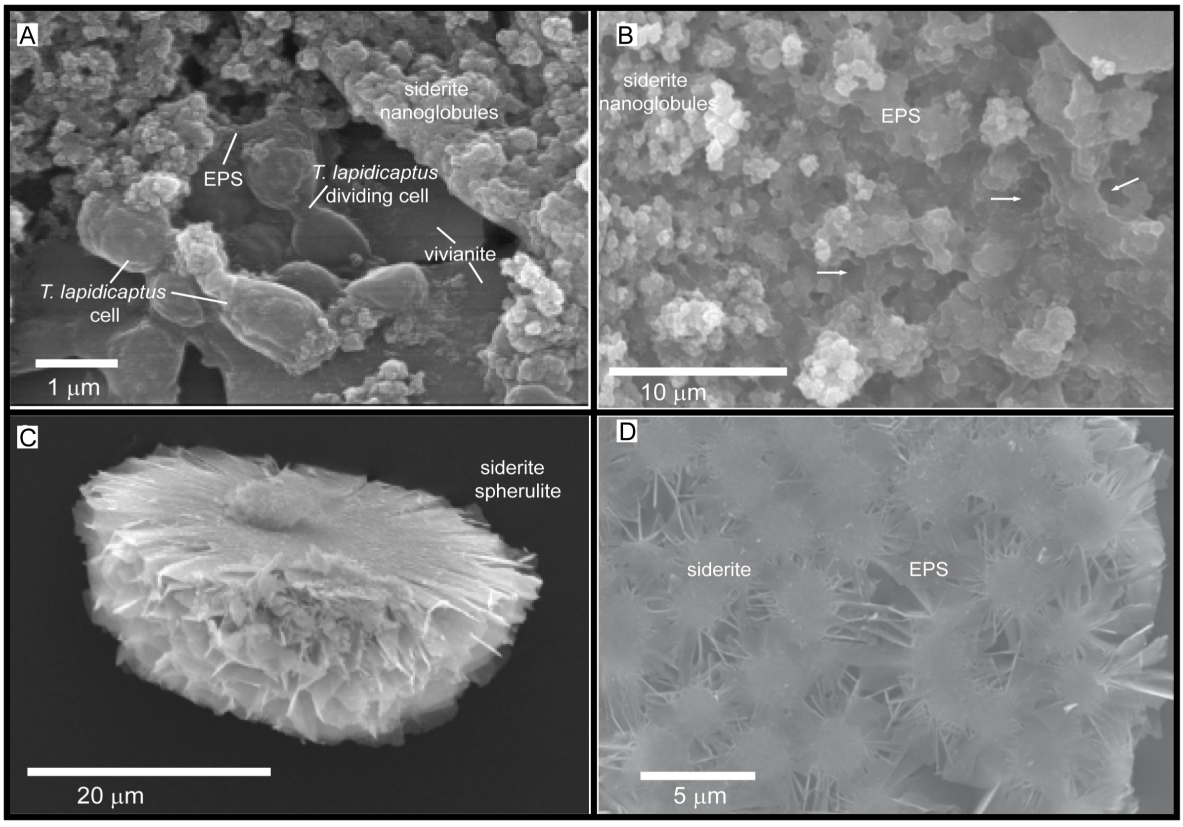


Figure 6.7. SEM images of the Fe-carbonate precipitates from IPBSL-7^T (*Tessaracoccus lapidicaptus)* anaerobic cultures. A) Siderite nanoglobules embedded in EPS and attached to mineralized IPBSL-7^T cells. Note the vivianite crystal below cells. **B)** Fe-carbonate nanoglobules (siderite) embedded in EPS and delimiting the bacterial cell contours (white arrows). These nanostructures display granulated texture. White arrows correspond to moulds of degraded bacteria (broken cells). **C)** Broken microspherulite of siderite. **D)** detail of a siderite spherulite which formed by aggregation of nanoparticles.

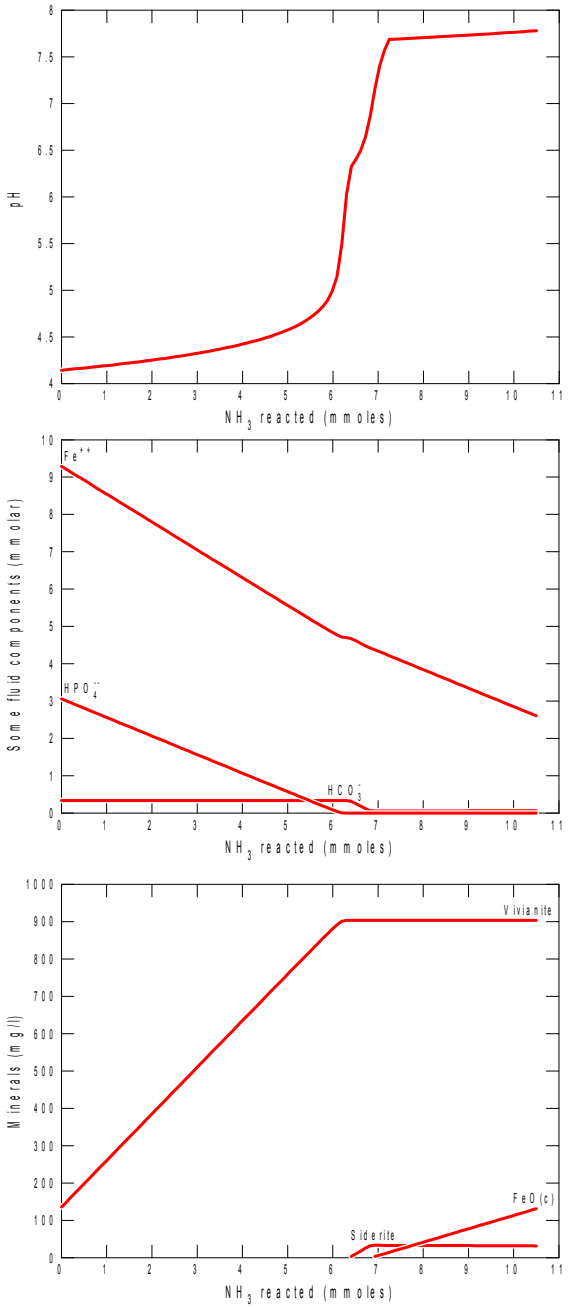


Figure 6.8. Geochemical modelling of mineral precipitation during bacterial growth in anoxic FE medium. The observed increase in pH in the inoculated cultures was simulated by incremental additions of NH_3 to the system. **(a)** pH evolution with NH_3 increase. **(b, d)** Evolution of aqueous species in solution (**b**) and precipitated minerals (**d**) with increasing NH_3 concentration. **(c, e)** Evolution of species in solution (**b**) and precipitated minerals (**d**) with increasing pH.

6.3. Discussion

6.3.1. Assignment of strain IPBSL-7^T to the *Tessaracoccus* genus

Phylogenetic analyses showed that strain IPBSL-7^T clustered together with other strains from the genus *Tessaracoccus* (**Figure 6.3**). Additionally, the strain was positive for nitrate reduction and its peptidoglycan type was A3γ' (Schelifer & Seidl, 1985). These two characteristics are shared among all known members of the genus *Tessaracoccus*, but are absent in closely related genera such as

Brooklawnia (Bae *et al.*, 2006) and *Propionimicrobium* (Stackebrandt *et al.*, 2002). This further supports the inclusion of strain IPBSL-7^T in the genus *Tessaracoccus*. The main morphological, physiological and chemotaxonomic characteristics that differentiate strain IPBSL-7^T from other related strains representing the genus *Tessaracoccus* are shown in **Tables 6.1** and **6.2**. There were differences in salt tolerance, substrate utilization, enzymatic activities and membrane composition. Interestingly, diphosphatidylglycerol, which is present in the previously described species of the genus *Tessaracoccus*, was not detected in strain IPBSL-7^T (**Figure 6.2**). This feature could be diagnostic for distinguishing strain IPBSL-7^T from other related strains. On the basis of these phenotypic and phylogenetic results, strain IPBSL-7 T was included in the genus *Tessaracoccus* as a representative of a novel species, for which the name *Tessaracoccus lapidicaptus* sp. nov. was proposed.

6.3.2. Description of *Tessaracoccus lapidicaptus sp. nov.*

Tessaracoccus lapidicaptus (la.pi.di.cap'tus. L. n. *lapis lapidis* stone; L. adj *captus* captive; N.L. masc. adj. *Lapidicaptus* stone-captive).

Facultatively anaerobic, Gram-positive, oxidase-negative, catalase-positive, non-endospore-forming, non-motile, oval to rod-shaped cells (0.45x0.5–1.0 μm) that often occur in pairs or small groups (**Figure 6.1**). Colonies grown on TSA are circular, convex, smooth and colourless. Growth occurs between 15 and 40 ºC, with an optimum at 37 ºC. The pH range for growth is from pH 6 to 9, with pH 8 being the optimal value. Tolerates up to 2 % (w/v) NaCl, 1 mM Zn^{2+} , 2 mM Ni^{2+} , 1 mM Cu^{2+} and 1 mM Co^{2+}. Is able to grow in the presence of guanidine hydrochloride, nalidixic acid, potassium tellurite and sodium butyrate; but not in the presence of 1 % sodium lactate, fusidic acid, D-serine, niaproof 4, tetrazolium violet, tetrazolium blue, lithium bromide or sodium bromate. Utilizes maltose, trehalose, cellobiose, sucrose, turanose, stachyose, raffinose, α-lactose, melibiose, D-salicin, α-D-glucose, D-mannose, D- fructose, D-galactose, D-sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol and acetoacetic acid. Cannot use dextrin, gentiobiose, N-acetyl-D-glucosamine, N-acetyl-β-D- mannosamine, N-acetyl-D-galactosamine, N-acetyl-D-neuraminic acid, 3-methyl glucose, D-fucose, L-fucose, L- rhamnose, inosine, D -glucose 6-phosphate, D-fructose 6-phosphate, D-aspartic acid, D-serine, gelatin, glycyl-L-proline, L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L- histidine, L-pyroglutamic acid, L-serine, pectin, D-galacturonic acid, D-galacturonic acid lactone, D-gluconic acid, D-glucuronic acid, glucuronamide, mucic acid, quinic acid, D- saccharic acid, *p*-hydroxyphenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, L-

lactic acid, citric acid, α-ketoglutaric acid, D-malic acid, L-malic acid, bromosuccinic acid, Tween 40, γ-aminobutyric acid, α-hydroxybutyric acid, β-hydroxy-DL-butyric acid, α-ketobutyric acid, propionic acid, acetic acid and formic acid. Produces acid from the following substrates: D-glucose, D-mannitol, glycerol, cellobiose, D-mannose, raffinose, D-rhamnose. Hydrolyses aesculin, but not gelatin. Is not able to produce indole from L-tryptophan and is urease-negative. No anaerobic acid production is detected from the following substrates: lactose, sucrose, maltose, salicin, D-xylose, D-arabinose, melezitose, D-sorbitol, trehalose. Positive for the following enzymic activities: esterase, esterase lipase, leucine arylamidase, α- galactosidase, β-galactosidase, α -glucosidase and β-glucosidase. No activity from the following enzymes was detected: alkaline phosphatase, lipase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsine, acid phosphatase, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, N- acetyl-β-glucosaminidase, α-manosidase and α-fucosidase. The peptidoglycan was of the A3γ' type and contained LL-DAP. The polar lipid composition includes phosphatidylglycerol, phosphatidylinositol, one phosphoglycolipid, two different phospholipids, three glycolipids and an unidentified polar lipid (**Figure 6.2**). The predominant cellular fatty acid is anteiso-$C_{15:0}$, but significant amounts of iso-$C_{15:0}$, iso-$C_{16:0}$ and $C_{16:0}$ are also found. More detailed results can be found in **Table 6.2**. The major respiratory quinones are MK-9(H_4), MK-9(H_6) and MK- 9(H_2). The type strain, IPBSL-7^T (=CECT 8385^T = DSM 27266^T), was isolated from a 297-metre-depth drilling core obtained from the Iberian Pyrite Belt. The DNA G+C content of the type strain is 70.3 mol%.

6.3.3. Bioprecipitation of iron phosphates and carbonates by *Tessaracoccus lapidicaptus*

Microbial mediated mineral precipitation has been previously reported for culture experiments using sulphate reducing-bacteria, aerobic heterotrophic bacteria, acidophilic iron-reducing bacteria and fungi (Aloisi *et al.*, 2006; Bontognali *et al.*, 2008; Sánchez-Román *et al.*, 2008, 2014; Oggerin *et al.*, 2013). Our TEM and SEM studies showed that carbonate and phosphate nanocrystals nucleated on bacterial cell surfaces and EPS (**Figure 6.5**). *T. lapidicaptus* cells were also found attached to the surface of microcrystals (**Figures 6.6; 6.7**), which in the case of vivianite reached the millimetre scale (**Figure 6.6c**).

Our simulations predicted that vivianite should have precipitated abiotically in the FE medium and the pH should have dropped to around four (**Figure 6.8a,b**, 0 mmol of NH_3 reacted). Instead, we saw no precipitation and no pH changes in neither of the experimental controls (uninoculated and

inoculated with autoclaved cells). It must be noted, however, that the software we used to model our culture medium assumes that the equilibrium is reached on every step of the simulation. This equilibrium may not have been reached during the course of our experiment. In fact, the saturation of mineral phases does not directly imply the abiotic precipitation of those minerals, since the process will depend on their precipitation kinetics (Morse, 1983). The lack of minerals on our controls indeed suggests that the rate of abiotic precipitation at the initial pH was too slow for it to develop during the course of our experiment.

Since vivianite precipitation was associated to a pH drop, the observed increase in pH (from 6 to 7.5) in our inoculated cultures must be the attributed to microbial activity. In order to account for that in our calculations, increasing concentrations of NH_3 (used here as a proxy for biotic basification of the culture medium) were added to the simulation. The results of the simulation were consistent with the results observed in our culture experiments, with vivianite (and to a lower extent) siderite having precipitated after the solution reached a pH of 7.5. This pattern (vivianite precipitating first, and in greater amounts) can be attributed to PO_4^{3-} inhibiting the precipitation of carbonate minerals (Bouropoulos & Koutsoukos, 2000; Morse *et al.*, 2007), and siderite having a lower stability at low pHs (Lemos *et al.*, 2007). On the other hand, the predicted precipitation of FeO at higher pHs can be disregarded, as FeO is thermodynamically unstable below 575ºC (Greenwood & Earnshaw, 2012).

Bacteria induce mineral precipitation by concentrating ions (e.g., Ca, Fe, Mg, CO_3^{2-}, PO_3^{4-}, NH_4^+) and changing the pH in the microenvironment surrounding their cells (Ehrlich & Newman, 2008; Van Lith *et al.*, 2003; Sánchez-Román *et al.*, 2011). Bacterial cells act as a template for mineral nucleation by adsorbing ions around the cellular surface membrane or cell wall (Schultze-Lam *et al.*, 1996; Bosak & Newman, 2003). Moreover, EPS are considered as an important factor for mineral precipitation (Dupraz *et al.*, 2009; Aloisi *et al.*, 2006; Ercole *et al.*, 2007; Bontognali *et al.*, 2008; Krause *et al.*, 2012). The high crystallinity of the vivianite crystals observed in our experiment (**Figure 6.4**) suggests that they are mostly of abiotic origin. Given the lack of crystals in the sterile controls, we propose that *T. lapidicaptus* cells promote the precipitation of vivianite by nucleating nanocrystals in their cell walls and associated EPS, which can then act as templates for abiotic crystal growth. A similar mechanism can be proposed for the precipitation of siderite, although the siderite phase was generally amorphous, which suggests a predominance of biotic over abiotic growth. Nanocrystals of both minerals were found attached to *T. lapidicaptus* cells by TEM-EDX imaging. The rise in pH produced by cell metabolism is another key factor for the precipitation of both minerals, especially siderite. *T. lapidicaptus* cells likely increased the pH of the

culture medium via amino acid fermentation to NH_3. Amino acids were present in the form of yeast extract (0.1 g/l) with a 6% of amino nitrogen. Even if all that nitrogen was fermented, the produced amount of NH_3 (roughly 0.35 mM) would be insufficient to raise the pH of the culture medium to 7.5, according to our simulation. However, and as stated before, the simulation assumes that thermodynamic equilibrium has been reached after each step, while in reality equilibrium was probably not reached during the course of our experiment. An incomplete precipitation of vivianite (which would imply a smaller acidification of the culture medium) can explain the observed results by allowing pH 7.5 to be reached with less NH_3 production.

The ability of *T. lapidicaptus* to remove considerable amounts of iron and phosphate might make it a good candidate for phosphorous removal in anaerobic systems. Furthermore, vivianite and siderite are good phosphorous and iron fertilizers (Eynard *et al.*, 1992; Sánchez-Alcalá *et al.*, 2012). This suggest that strain *T. lapidicaptus* might have industrial applications, although their exploration is beyond the scope of this thesis.

6.3.4. Potential roles of *Tessaracoccus lapidicaptus* in the deep IPB subsurface

Two members of the *Tessaracoccus* genus have been previously isolated from marine sediments (Lee & Lee, 2008) and deep subsurface environments (Finster *et al.*, 2009). Others have been isolated from crude oil-contaminated saline soil (Cai *et al.*, 2011) and oleaginous, water-mixed metalworking fluids (Kämpfer *et al.*, 2009), which suggests that the *Tessaracoccus* genus might be especially adept at degrading hydrocarbons and/or recalcitrant organic matter under harsh environmental conditions. This might also be the case of *Tessaracoccus lapidicaptus*: even though the main objective of the physiological characterisation reported in this chapter was the description of *T. lapidicaptus* as a novel *Tessaracoccus* species, our results showed that it is able to degrade complex organic matter to acetate and reduce nitrate (**section 6.2.1**, **Table 6.1**), whilst withstanding moderate concentrations of heavy metals. In the IPB subsurface, *T. lapidicaptus* might be degrading recalcitrant organic matter and making it available to other members of the microbial community. The salt tolerance of strain IPBSL-7^T was of only 2% (w/v), which is a somewhat low value. However, salt resistance is not a key feature to survive in deep subsurface environments, in which desiccation is a more likely source of stress (see **section 1.1.4**).

The results of our biomineralization experiments indicate that *T. lapidicaptus* might be able to influence the mineral composition of its bulk environment, and to accumulate nutrients such as iron

and phosphate via local precipitation in their cell walls and surrounding EPS. Biomineralization has also been suggested to increase desiccation resistance (Phoenix & Konhauser, 2008). No vivianite was detected in borehole BH11, from which *T. lapidicaptus* was isolated, which may be explained due to the presence of dissolved sulphides inhibiting vivianite precipitation (Postma, 1981; Manning *et al.*, 1999). On the other hand, siderite was found in the BH11, albeit at trace amounts (Mónica Sánchez-Román, personal communication). However, significant quantities of biogenic siderite have been detected in shallower regions of the IPB (Fernández-Remolar *et al.*, 2012). While vivianite and siderite were the only minerals precipitated by *T. lapidicaptus* in our experiment, it might be capable of precipitating other minerals under different experimental (or environmental) conditions.

While the results presented in these chapter allowed us to hypothesize about the role of *T. lapidicaptus* in subsurface ecosystems, its actual relevance (or lack thereof) in the IPB environment can not be assessed by enrichment cultures alone (see **section 1.4.5**). Molecular analyses would be needed to determine whether *T. lapidicaptus* is actually abundant in our samples (see next chapter). Additionally, more studies, including the sequencing of its genome, would help to fully elucidate its adaptation mechanisms to deep subsurface environments.

7. THE SIGNAL-TO-NOISE PROBLEM IN DEEP SUBSURFACE METAGENOMICS: WET-LAB AND COMPUTATIONAL SOLUTIONS

7.1. Introduction

7.1.1. The signal-to-noise problem in deep subsurface metagenomics

The isolation and sequencing of environmental DNA has become a standard practice for studying ecosystems, especially after the appearance of high-throughput sequencing techniques. By studying the genetic information present in a given microbial community, or metagenome, we can infer not only its taxonomic composition, but also its metabolic potential. Metagenomic analyses have been successfully applied to many environments (see Yoon *et al.*, 2015; Brum & Sullivan, 2015; de Castro *et al.*, 2014; Prussin *et al.*, 2014; van Baarlen *et al.*, 2013; Mendes *et al.*, 2013 for recent reviews). However, and to the best of our knowledge, only a few of those works have been conducted on deep subsurface ecosystems (Mason *et al.*, 2008; Chivian *et al*. 2008; Osburn *et al.*, 2014). Suitable deep subsurface samples are not only difficult to obtain (see **section 1.4.1**), but are also one of the most challenging materials on which to conduct metagenomic analyses. Energy, space and water limitations result in very little to no microbial biomass, while the presence of a complex mineral matrix can greatly reduce the yield of nucleic acid extraction protocols via cellular and molecular adsorption and oxidative damage (see **section 1.4.6.3**). On the other hand, a certain degree of contamination is expected in nucleic acid based studies, coming from non-aseptic sampling (with the fluids used to cool the drill bit and stabilizing the borehole being a common contaminant in direct deep subsurface sampling projects), laboratory mispractice or even supposedly sterile commercial solutions and kits (Salter *et al.*, 2014). Thus, the amount of true environmental nucleic acids we can obtain from deep subsurface samples (the *signal*) is severely low, while the amount of contaminant nucleic acids (the *noise*) remains the same as with other types of samples, or even increases. If extreme care is not taken, contaminants may mask the true environmental nucleic acids, resulting in erroneous depictions of the studied microbial communities.

7.1.2. Increasing the signal: Improved DNA extraction and amplification

The detrimental effects of mineral matrices (such as the silica-rich BH10 samples, see **sections 1.3.1** and **3.4.3**) in nucleic acid extraction can be partially overcome by choosing a suitable extraction protocol. Indirect approaches aim to first detach the microbial cells from the mineral and then perform the nucleic acid extraction, while direct approaches perform cellular lysis in the presence of the mineral, and include a chemical denaturation step in order to separate the DNA from the mineral matrix. In this work we have tested two different extraction protocols, one indirect (modified from Vilchez *et al.*, 2007) and one direct (based on Direito *et al.*, 2012), with mineral samples from borehole BH10 (see **sections 4.3.1** and **4.3.2**).

Even if the extraction is successful, the amount of recovered DNA will likely be insufficient for downstream applications. Multiple displacement amplification (MDA) remains the method of choice for obtaining sizeable amounts of DNA from a minute quantity of starting template, even single DNA molecules (Braslavsky *et al.*, 2003). This is however a double-edged sword, as even trace amounts of contaminants will be also amplified. Additionally, MDA is known to introduce amplification biases, compromising the quantitative interpretation of the obtained results (Yilmaz *et al.*, 2010). These two problems must be taken into account when applying MDA to low-biomass environmental samples (see next sections).

7.1.3. Reducing the noise: Minimizing the effect of contaminants during DNA extraction and amplification

As discussed above, deep subsurface metagenomic studies are especially susceptible to sample and laboratory contamination issues. The main challenges and solutions for aseptic sample retrieval have been introduced in **section 1.4.1**, while laboratory contaminations can be avoided by using standard aseptic techniques.

Commercial solutions and kits can also contain contaminant DNA (Salter *et al.*, 2014), a fact that can be critical during the DNA extraction and amplification step. Therefore, all solutions used during these two steps must be considered as potentially contaminated, and treated likewise. Contaminant DNA can be easily removed from inorganic solutions by UV light treatment. However, MDA requires the use of primers, nucleotides and a polymerase, which are themselves sensitive to UV light. In this work we have optimized a UV decontamination protocol based on the work from Woyke *et al.*, (2011) that efficiently removes genomic DNA from MDA reagents without impairing the DNA-polymerase activity of the resulting reaction mix.

7.1.4. Reducing the noise: Minimizing the effect of sequencing errors in 16S community profiling

High-throughput sequencing of marker genes, such as the 16S ribosomal RNA, has become an invaluable tool for microbial ecologists, since it allows for a previously unreachable level of detail in the analysis of complex microbial communities. Many studies have used platforms such as the Roche 454, Illumina or IonTorrent sequencers to thoroughly characterise and compare microbial communities at an affordable cost (Roesch *et al.*, 2007; Costello *et al.*, 2009; Caporaso *et al.*, 2011; Yatsusenko *et al.*, 2012; Ding & Schloss, 2014), while others have taken advantage of their very high yield in order to analyse the structure and composition of the rare biosphere (Hugoni *et al.*, 2013). However, the correct assessment of sequencing artefacts is critical in obtaining representative results. In whole genome sequencing studies an erroneous base can be corrected by overlapping reads during consensus sequence assemblage, but in marker-gene studies each read is assumed to come from a different individual in the community. In this case, sequencing errors can cause the mis-clustering of otherwise similar reads, resulting in the overestimation of microbial diversity (Kunin *et al.*, 2010). The most common software tools and packages include sequence clustering into OTUs in their recommended pipelines (Schloss *et al.*, 2009; Caporaso *et al.*, 2010; Edgar, 2010; Huang *et al.*, 2010; Quince *et al.*, 2011; Schloss et al 2011; Caporaso *et al.*, 2012; see Edgar, 2013 for a comparison of several molecular ecology pipelines). Alternatives to traditional clustering have been recently proposed, such as distribution-based clustering (Preheim *et al.*, 2013) or a clustering-free approach (Tikhonov *et al.*, 2015). These novel methods are especially suited for subpopulation level studies, but work only for moderate-to-high abundance sequences, being unsuitable for population-level alpha or beta diversity studies. Moreover, even although they can remove likely-erroneous sequences and resolve subpopulations based on dynamic information, they nevertheless rely on a quality-filtering step for the pre-processing of raw reads.

Amplicon denoising (Quince *et al.*, 2011, Reeder & Knight, 2010) is a widespread method for filtering Roche 454 pyrosequencing reads that can also be applied to IonTorrent data. It works on flowgrams rather than sequences, which allows for a more natural modelling of the homopolymer read errors that are characteristic of pyrosequencing and ion semiconductor sequencing. However, it is platform specific and computationally expensive.

For Illumina systems there is no consensus approach to quality-filtering, with the authors of mothur (Kozich *et al.*, 2013), QIIME (Bokulich *et al.*, 2013) and UPARSE (Edgar, 2013) proposing different solutions. All those heuristic approaches were published as parts of their respective

pipelines, but to the best of our knowledge they have not been thoroughly compared to each other.

The lack of a rigorous method for incorporating quality scores in the analysis of marker-gene sequences has also led some authors to advocate for a stringent filtering in order to reduce the retrieval of spurious diversity (Caporaso *et al.*, 2011). However, over-stringent filtration will result in an undesired loss of sensitivity and will have an impact on the observed taxonomic distribution (Bokulich *et al.*, 2013). Therefore, an accurate algorithm that overcomes these problems is desirable.

Herein we present and validate the Poisson binomial filtering (PBF) method, which is able to determine the exact error-probability distribution of any sequence with associated quality scores, by using a simple statistical approach. Phred quality scores, which represent the probability that a given base call is mistaken, can be derived from the raw output of every sequencing platform. Reading a single base can be likened to tossing a coin: the base is either right or wrong, and both chances can be determined from its quality score. In fact, the number of errors present in a given base follows a Bernoulli distribution, i.e. a binomial distribution with a single trial. For a sequence of nucleotides with different error probabilities, we sum their associated Bernoulli random variables in order to obtain the exact probability that the sequence has accumulated more than k errors, where k is the maximum number of errors that still allows for a correct clustering (see **section 4.3.4; Annex 2 - Supplementary Note 1**). This algorithm is based on simple statistical principles and, since it only requires Phred quality scores as an input, it is expected to work robustly regardless of the sequencing platform. It is also computationally efficient, scales linearly with the number of sequences, and has a low memory fingerprint, making it useful even in low-performance desktop environments.

7.1.5. Additional safeguards: Post-sequencing assessment of contaminants and MDA amplification biases

Even with optimised extraction and amplification protocols, contamination with exogenous DNA can not be discarded. In order to measure the degree of contamination of our MDA-amplified DNA samples, we pyrosequenced the 16S rRNA genes of five subsurface samples and compared the resulting taxonomic profiles with that of a pool of drilling water samples collected during the IPBSL campaign. The presence of drilling water-related OTUs was used as a tracer for sample contamination during drilling. Additionally, all the OTUs were classified according to their most probable origin (see **section 4.3.14)** and the presence of animal microbiome-related OTUs (which

would suggest contamination during sample manipulation) was assessed. In the final analyses (see next chapter) a DNA extraction blank was also included and subjected to MDA and 16S rRNA gene sequencing, in order to fully account for the whole DNA extraction and amplification procedure.

For two of the five samples, we were also able to directly amplify the 16S rRNA gene without the need for an intermediate MDA step. By comparing the taxonomic profiles of these two samples with and without MDA, we were able to assess the degree of taxonomic bias introduced during the amplification step.

7.2. Results

7.2.1. Comparison between the direct and indirect DNA extraction protocols

Since in all cases the DNA extraction yield was below the detection limit, efficiency was measured as the ability to produce extracts amplifiable by direct PCR of the 16S rRNA gene. Overall, the indirect approach was successful in 7 out of the 28 (25%) tested samples, while the direct approach was successful in 7 out of the 12 (58%) samples. Five samples were extracted with the two protocols (**Table 7.1**), with the indirect approach succeeding in only one case and the direct approach succeeding in three cases. Based on these results, the direct DNA extraction approach proposed by Direito *et al.*, (2012) was chosen as the method of choice for further metagenomic analyses of the BH10 borehole.

Table 7.1. Comparison of indirect and direct protocols for isolating DNA from deep subsurface samples. Since in all cases the DNA extraction yield was below the detection limit, efficiency was measured as the ability to produce extracts amplifiable by direct PCR of the 16S rRNA gene.

Sample (mbsf)	Indirect extraction / 16S PCR	Direct extraction / 16S PCR
336	-	+
355	-	-
392	-	+
420	+	+
544	-	-

7.2.2. An optimized protocol for MDA reagent decontamination

MDA reagents spiked with 0.05 ng of genomic DNA from *Leptospirillum ferroxidans* were subjected to increasingly longer UV treatments as described in Woyke *et al.*, (2011). After UV treatment, a set of tubes was spiked again with another 0.05 ng of genomic DNA in order to determine whether the treated mix was still capable of amplifying DNA, while a second set of tubes was left unspiked to determine whether the original contamination had been removed by the treatment. Finally, the tubes were subjected to MDA and 16S rRNA gene PCR as described in **section 4.3.3.**

UV-treated mixes retained their polymerase activity along the whole range of energies tested, yielding sizeable amounts of product in all cases (**Figure 7.1A,B**). It must be noted, however, that MDA is known to generate high molecular weight products even in the absence of a template, due to spurious self-priming of random hexamers (Woyke *et al.*, 2011). Considering this, we used 16S rRNA gene PCR in order to further confirm that the amplified DNA was indeed of genomic origin. All the re-spiked tubes showed positive 16S amplification, but no 16S amplification was detected in the tubes that were not re-spiked (**Figure 7.1C,D)**. This shows that a UV treatment of at least 0.02 mJ/cm^2 was enough to remove trace amounts of contaminant DNA without affecting the ability to amplify the same trace amounts of environmental DNA. Based on these results, an UV energy of 0.04 mJ/cm^2 was applied to decontaminate MDA reagents in further experiments.

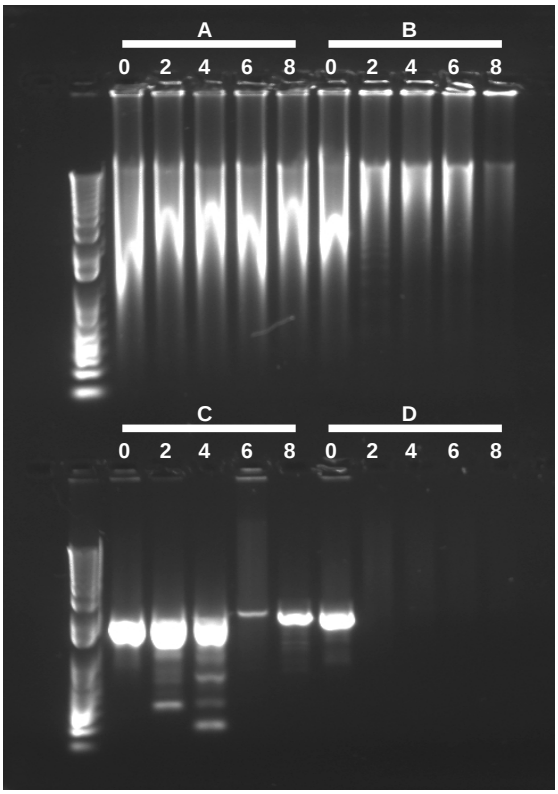


Figure 7.1. Removal of contaminant genomic DNA from MDA reagents. MDA reagents were spiked with a minute amount of *Leptospirillum ferroxidans* genomic DNA and subjected to increasingly long UV light treatments.

Agarose gel showing: (**A**) MDA with reagents spiked again with genomic DNA after UV treatment; (**B**) MDA with UV-treated reagents without re-spiking; **(C**) 16S rRNA gene PCR of the MDA products shown in **A**; **(D**) 16S rRNA gene PCR of the MDA products shown in **B**; Leftmost wells: Invitrogen 1Kb Plus DNA ladder. The numbers below **A, B, C** and **D** correspond to the UV energy (expressed in 10^{-2} mJ/cm^2) used to treat each MDA reagent mix.

7.2.3. Validation of the Poisson binomial filtering algorithm

We validated the Poisson binomial filtering algorithm and compared it with the different filtering approaches recommended by the authors of mothur (Schloss *et al.*, 2009; Schloss *et al.*, 2011; Kozich *et al.*, 2013), USEARCH-UPARSE (Edgar, 2010; Edgar, 2013) and QIIME (Bokulich *et al.*, 2013) by quality-filtering datasets obtained by sequencing different mock and environmental microbial communities with the Roche 454 GS FLX Titanium, the Illumina MiSeq and the IonTorrent PGM platforms. In order to evaluate the different methods on equal grounds, filtered reads were processed with a common downstream pipeline that included chimera-filtering with UCHIME (Edgar *et al.*, 2011), sample size standardization and OTU clustering.

7.2.3.1. Poisson binomial filtering accurately discriminates between good and erroneous sequences

When applying our default cut-off of 1% errors allowed per sequence, our algorithm accurately classified 96% of the mock community sequences from the Even1M dataset (**Figure 7.2d**). 3% of the sequences were incorrectly discarded while, remarkably, only 1% of the sequences were incorrectly retained. Moreover, most of those incorrectly retained sequences had only 3 true errors (1.2% errors per sequence), meaning that they would likely cluster correctly when applying the standard 3% OTU distance cut-off. The rest of the Illumina datasets rendered similar results. The accuracy of our method was slightly lower for the 454 and IonTorrent datasets, but it nevertheless resulted in a minimum of 88% (for 454) and 79% (for IonTorrent) correctly classified sequences. (**Annex 2 - Supplementary Figure SN4.1**).

7.2.3.2. Performance of the different filtering methods on mock community datasets

Publicly available datasets from even and staggered mock communities from the Human Microbiome Project (Haas *et al.*, 2011) were filtered with PBF, mothur, USEARCH and QIIME (**Figure 7.3**; **Annex 2 - Supplementary Note 4**). These artificial communities contain known amounts of 16S rRNA gene copies from 20 different bacterial organisms. The fact that both the qualitative and quantitative composition of the samples are known beforehand allowed us to thoroughly compare the effects of the different filtering methods in terms of OTU accuracy, alpha diversity and community composition. OTU accuracy was defined as the maximum similarity of its representative sequence to the 16S sequences of the microorganisms used to build the mock community, as previously described in Edgar, (2012). We were also interested in determining how

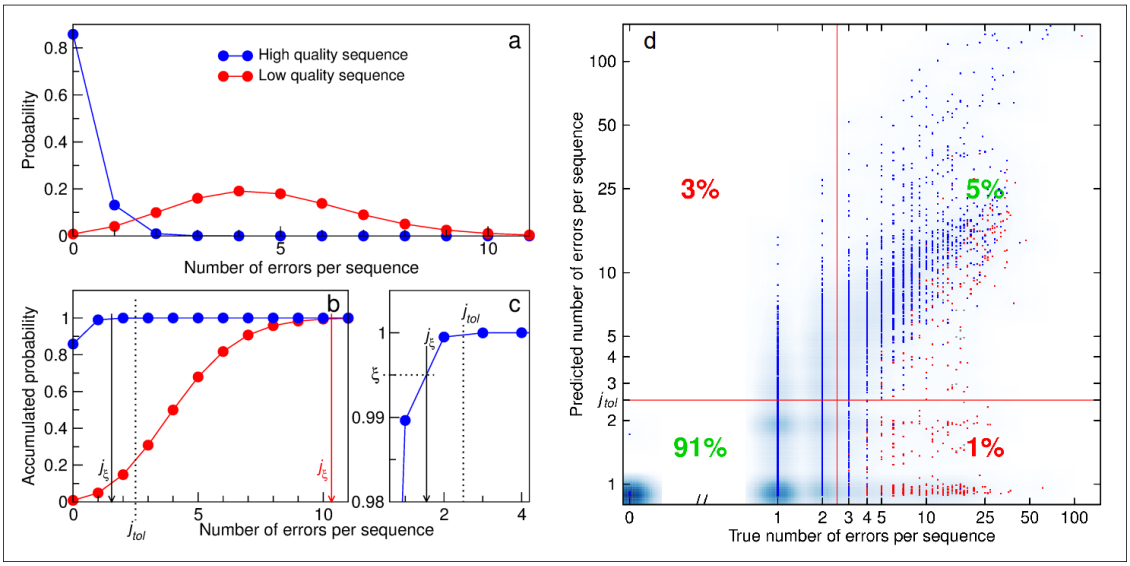


Figure 7.2. Poisson binomial filtering accurately discriminates between good and erroneous sequences.
(**a, b**): Error probability distribution (**a**) and accumulated error probability distribution (**b**) of two example nucleotide sequences, as calculated from their quality scores by the Poisson binomial filtering algorithm. j_ξ stands for the 99.5th percentile of the error probability (i.e. a sequence has a probability $\xi = 0.995$ of having less than j_ξ errors). j_{tol} is the maximum tolerable number of errors (1% of the sequence length in our case). Sequences with $j_\xi > j_{tol}$ are discarded in the filtering step.
(**c**) Zoom of (**b**), sketching the calculation of j_ξ for the high quality sequence.
(**d**) Comparison between the number of errors j_ξ predicted by the Poisson binomial algorithm and the true number of errors for all sequences from the Even1T mock community dataset (see **Annex 2 - Supplementary Note SN3**). Dots represent unique sequences. True mock community sequences are plotted in blue, contaminant sequences are plotted in gray, and chimeric sequences are plotted in red. The blue background represents sequence abundance (note that few unique sequences may have a high number of representatives, and vice versa). Red lines indicate our error cut-off of 2.5 errors per sequence (j_{tol}). The plot is thus divided in four quadrants corresponding to correctly retained sequences (lower left), correctly discarded sequences (upper right), incorrectly discarded sequences (upper left) and incorrectly retained sequences (lower right). The percentage of true mock community sequences present on each quadrant is also indicated. Poisson binomial filtering correctly classified 96% of the non-chimeric/non-contaminant sequences present in the Even1T dataset. The graph is plotted in logarithmic scale (the 0 in the x-axis is added for clarity).

the different filtering processes affected the observed community composition. The taxonomic bias in community composition caused by any given filtering method was calculated as the Bray-Curtis dissimilarity between the raw and the filtered datasets, after taxonomically classifying their reads down to the genus level.

In the even datasets, which contain the same number of 16S rRNA gene copies for each organism, all methods resulted in more than 20 OTUs after clustering. This was not surprising, since

contaminations, PCR errors and sequencing errors were expected to inflate the observed diversity. In the staggered communities, in which the number of 16S rRNA gene copies varied by several orders of magnitude between the different organisms, the observed diversity was generally lower, due to some species being present at very low abundances. The total number of reported OTUs

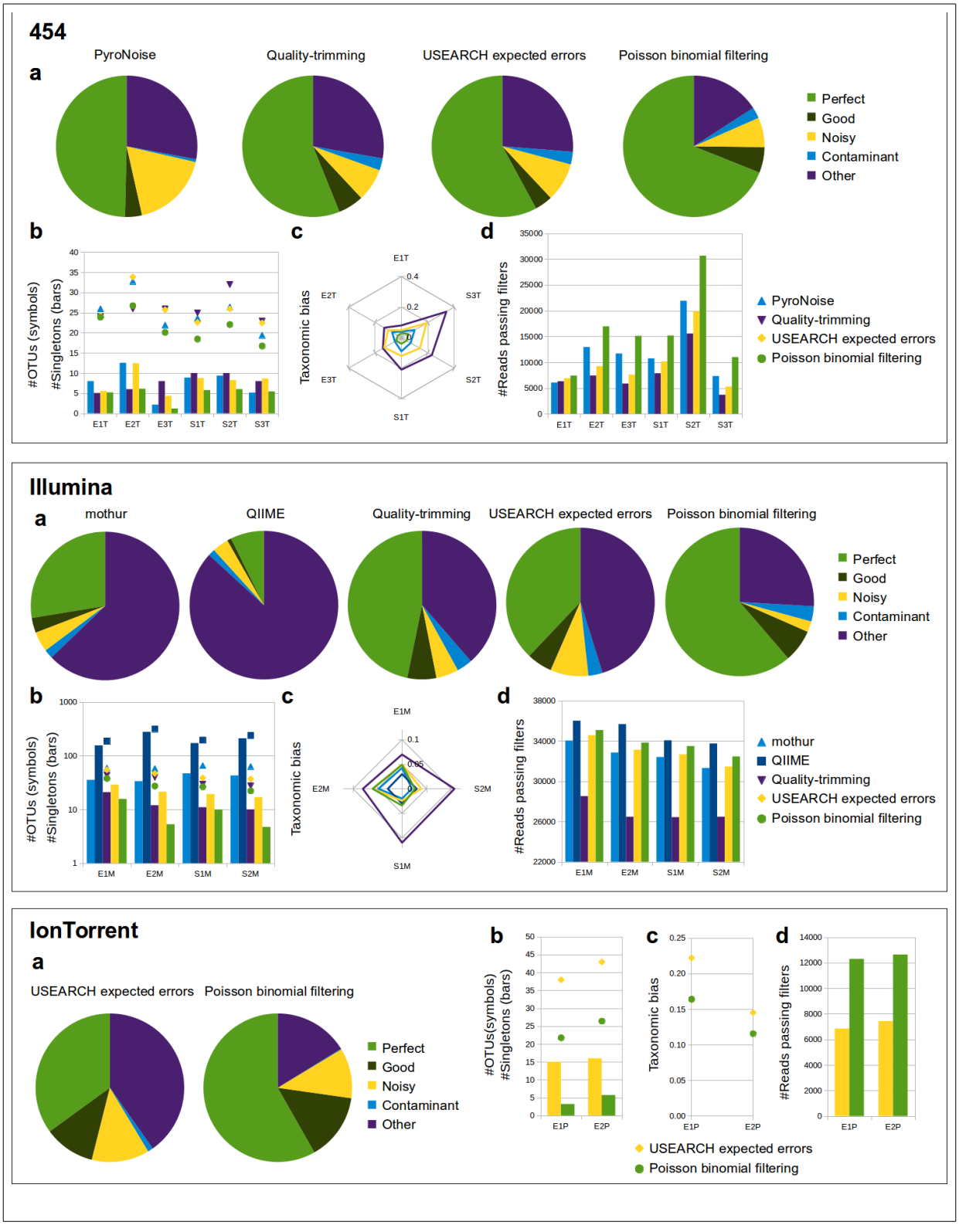


Figure 7.3. Comparison of filtering methods on 16S mock communities sequenced with the 454 GS FLX Titanium, Illumina MiSeq platforms and IonTorrent PGM platforms.
(**a**) Pie charts constructed by averaging the fraction of OTUs on each accuracy category along the six 454 or the four Illumina samples.
(**b , d**) Number of singletons (**b**, bars), total species (**b**, symbols) and reads (**d**) retrieved after filtering the raw reads with the different methods and performing chimera removal and clustering with a common pipeline. OTU and singleton numbers were obtained by averaging the results from 100 independent library size standardizations.
(c) Taxonomic bias caused by the different filtering methods, measured as the Bray-Curtis dissimilarity between the raw and the filtered read communities.

greatly varied between filtering methods, with Poisson binomial filtering consistently resulting in values that were the closest to the true diversity of the samples.

PBF also produced the highest proportion of accurate OTUs in all the 16S mock datasets for both sequencing platforms, while minimizing the number of singletons and spurious OTUs retrieved (**Figure 7.3a,b**). In the 454 and IonTorrent datasets it also discarded the smallest number of reads and resulted in the smallest taxonomic bias (**Figure 7.3c,d**). In the Illumina datasets QIIME retrieved a larger number of reads, while both QIIME and mothur caused smaller taxonomic biases than our method. (**Figure 7.3c,d - Illumina**). However, we believe that this was the result of a too shallow filtering by mothur and QIIME, since both methods produced a remarkably lower proportion of accurate OTUs and a larger number of OTUs and singletons (**Figure 7.3a,b - Illumina**). QIIME produced an especially high number of spurious OTUs, a fact that has also been discussed elsewhere (Edgar, 2013). Their pipeline (Bokulich *et al.*, 2013) deals with this problem by applying a post-hoc OTU size cut-off at the cost of sensitivity. Nonetheless, our results show that, even after the removal of singletons from the QIIME-filtered dataset, their number of OTUs would exceed that of the dataset filtered with our method, including singletons (**Annex 2 - Supplementary Table SN4.4**).

The two filtering algorithms included in the USEARCH suite showed an intermediate performance in terms of the number and accuracy of the OTUs retrieved for both the 454 and Illumina platforms. Quality trimming yielded the smallest number of reads and resulted in the highest taxonomic bias, which supports the idea that over-stringent filtering may lead to undesirable effects. In the IonTorrent datasets, USEARCH filtering performed below Poisson binomial filtering for all the studied benchmarks (**Figure 7.3 - IonTorrent**). Finally, the mothur implementation of the PyroNoise algorithm (Quince *et al.*, 2011) showed lower OTU accuracy than the other methods tested for filtering 454 reads. It has been previously described that the denoising process can introduce minor alterations in the original reads (Gaspar & Thomas, 2013), a phenomenon that

might explain these results. It must be noted that, albeit a pipeline for filtering IonTorrent reads with PyroNoise has been described, the IonTorrent mock community datasets were only available in Fastq format (Stephen Salipante, personnal communication), which precluded the use of flowgram denoising algorithms. However, this limitation was not present for the environmental datasets, and a comparison of quality filtering algorithms for IonTorrent datasets that includes PyroNoise can therefore be found in **Annex 2 - Supplementary Figure SN5.3**.

7.2.3.3. Performance of the different filtering methods on environmental datasets

The performance of the different filtering methods was also evaluated by quality-filtering publicly available datasets obtained by sequencing environmental communities (**Annex 2 - Supplementary Note 5**). The results were similar to the ones obtained with the mock communities, with Poisson binomial filtering being the most consistent method in producing the smallest number of OTUs and singletons. Additionally, the OTUs obtained with PBF were overall the most similar to the 16S sequences present in the SILVA 16S reference alignment (Quast *et al.*, 2012), which suggests that they contained the smallest number of errors. In the environmental 454 datasets, PyroNoise showed better results than in the 454 mock communities, but did it in an irregular fashion, especially in terms of OTU accuracy (**Annex 2 - Supplementary Figure SN5.1d**). This inconsistency may be again due to the alteration of the original reads, and suggests that PyroNoise requires a finer parameter optimization than other approaches in order to be fully effective. In the environmental IonTorrent datasets PyroNoise discarded the smallest number of reads, but resulted in the highest number of singletons and OTUs, which also borne the least similarity to the reference alignment. USEARCH showed an intermediate performance between PyroNoise and Poisson binomial filtering (**Annex 2 - Supplementary Figure SN5.3**). Finally, in the environmental Illumina datasets all filtering methods showed a similar behaviour to that in the mock communities (**Annex 2 - Supplementary Figure SN5.2**).

7.2.3.4. Quality-filtering is an additional source of taxonomic bias in microbial ecology studies

Even though the major sources of taxonomic biases in marker-gene-based studies are often related to PCR and library construction (Polz & Cavanaugh, 1998; Klindworth *et al.*, 2012; Ross *et al.*, 2013), the read filtering process can greatly exacerbate this problem (**Annex 2- Supplementary Note 2**).

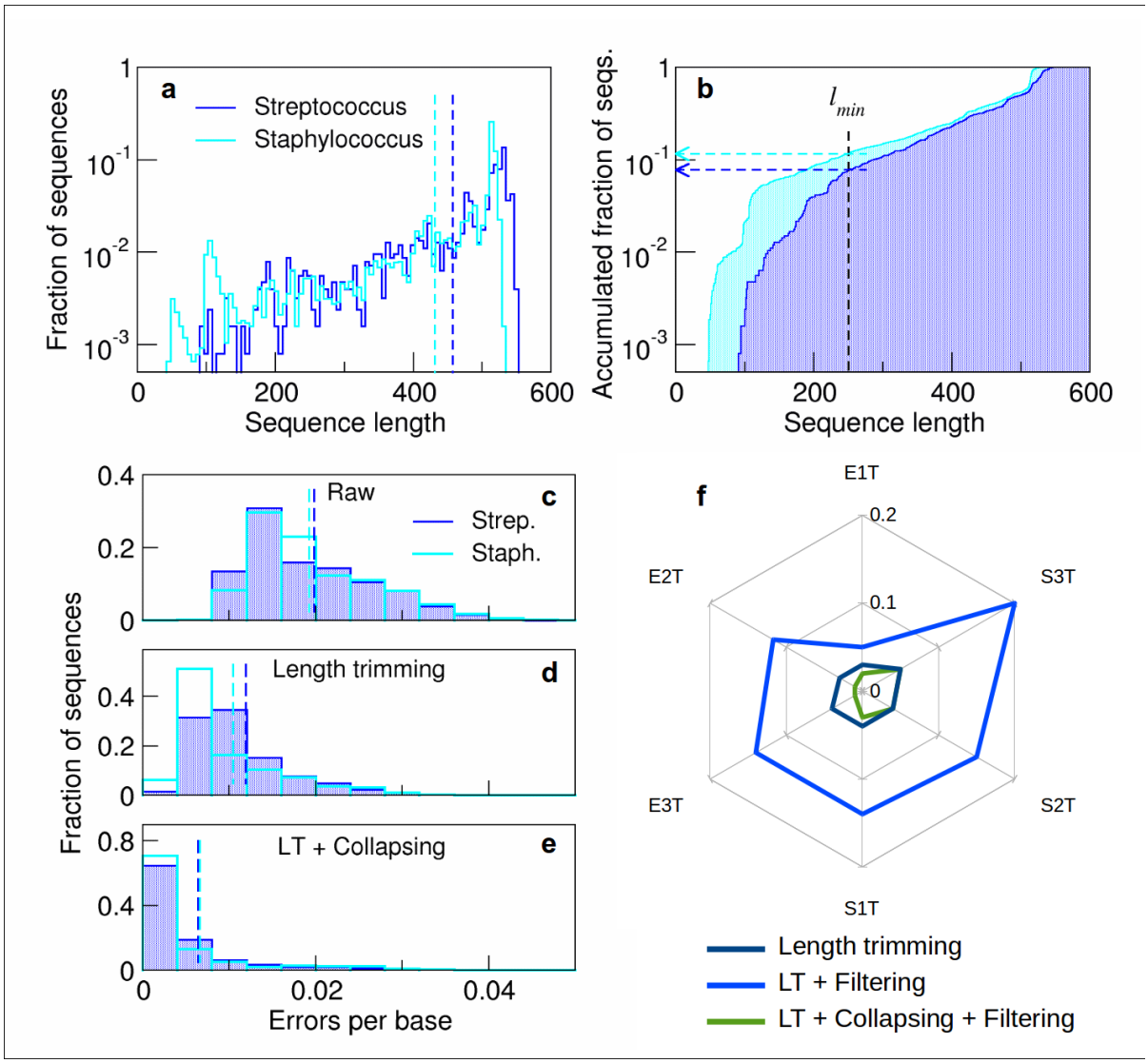


Figure 7.4. Addressing the taxonomic bias generated during the pre-processing and quality filtering of raw sequences.
(a, b): Raw reads from *Streptococcus* and *Staphylococcus*, the two most abundant genera in sample Even3T (see **Supplementary Note 3**), show different length distributions. The dashed vertical lines in **(a)** indicate the average read lengths. The arrows in **(b)** indicate the fraction of reads from each taxon removed after discarding sequences shorter than l_{min}= 250 nt.
(c, d, e): Errors per base distributions of *Streptococcus* and *Staphylococcus* reads in the **(c)** raw dataset, **(d)** after trimming the reads to 250 nt and discarding the ones shorter than the cut-off, and **(e)** after collapsing the trimmed reads. The dashed lines indicate average errors per base. Note that length trimming substantially increases the difference between the Streptococcus and Staphylococcus error distributions **(d)** when compared to that of the raw reads **(c)**. Filtering at this point would cause a 56.2% overrepresentation of *Streptococcus* versus *Staphylococcus* (see text and **Annex 2 - Supplementary Note 2**). Collapsing identical reads prior to filtering solves this problem **(e)**, reducing the overrepresentation to 1%.
(f): Compositional bias generated during the pre-processing and filtering of the six 454 mock community samples, measured as the Bray-Curtis dissimilarity between the raw and the processed read communities. This shows that results in (**c, d, e**) can be generalized to all the taxa present in all the samples.

We found significant biases in length and quality distribution between raw reads coming from different taxa in the mock 454 datasets (**Figure 7.4a,b,c**). Trimming them to a fixed length generated an artificial enrichment of the taxa with longer reads (**Figure 7.4b**), but since there is a decrease in quality at the end of 454 reads (see Edgar, 2013), it also resulted in a lower average read quality for the taxa with smaller raw reads (**Figure 7.4d**). This led to the generation of further taxonomic bias during the quality-filtering step (**Figure 7.4f**, **Annex 2 - Supplementary Note 2**). Similar biases have been previously found in IonTorrent reads (Salipante *et al.*, 2014), and were confirmed during this study (**Annex 2 - Supplementary Note 2)**. Biases in read quality distribution between different taxa were also found for the mock Illumina datasets, although to a lesser extent. We solved this problem by collapsing identical reads and choosing the one with the highest quality as a representative for filtering, in order to decide whether the whole group was discarded or allowed into the filtered dataset. This procedure reduced the effect of quality distribution biases, as even low abundance sequences are expected to have a high quality representative. Our solution rendered similar quality distributions for the different taxa, even after length trimming (**Figure 7.4e,f**), and significantly lower taxonomic biases than other filtering approaches, especially for 454 data (**Figure 7.3c**). Every method that relies on quality scores for sequence filtering will be affected by this source of bias. We therefore propose the approach described above as a general solution to this problem, since its simplicity makes it very easy to integrate into any filtering pipeline.

7.2.4. Post-sequencing assessment of contaminants and MDA amplification biases

Five powdered mineral samples from subsurface cores were subjected to DNA extraction, MDA and PCR of the 16S rRNA gene followed by pyrosequencing of the resulting amplicon. For two of those samples (420 and 607), the 16S rRNA gene was directly amplifiable from the DNA extract without the need for an intermediate MDA step. For those samples, 16S amplicons were generated and pyrosequenced for both the MDA and the non-MDA treated DNAs, in order to obtain an estimate of the taxonomic bias introduced during the MDA step. The resulting sequences were subjected to an standard bioinformatic pipeline and clustered into OTUs, which were then classified according to their most probable origin (see **section 4.3.14**). Finally, the samples were clustered based on their degree of similarity at the 0.09 OTU level.

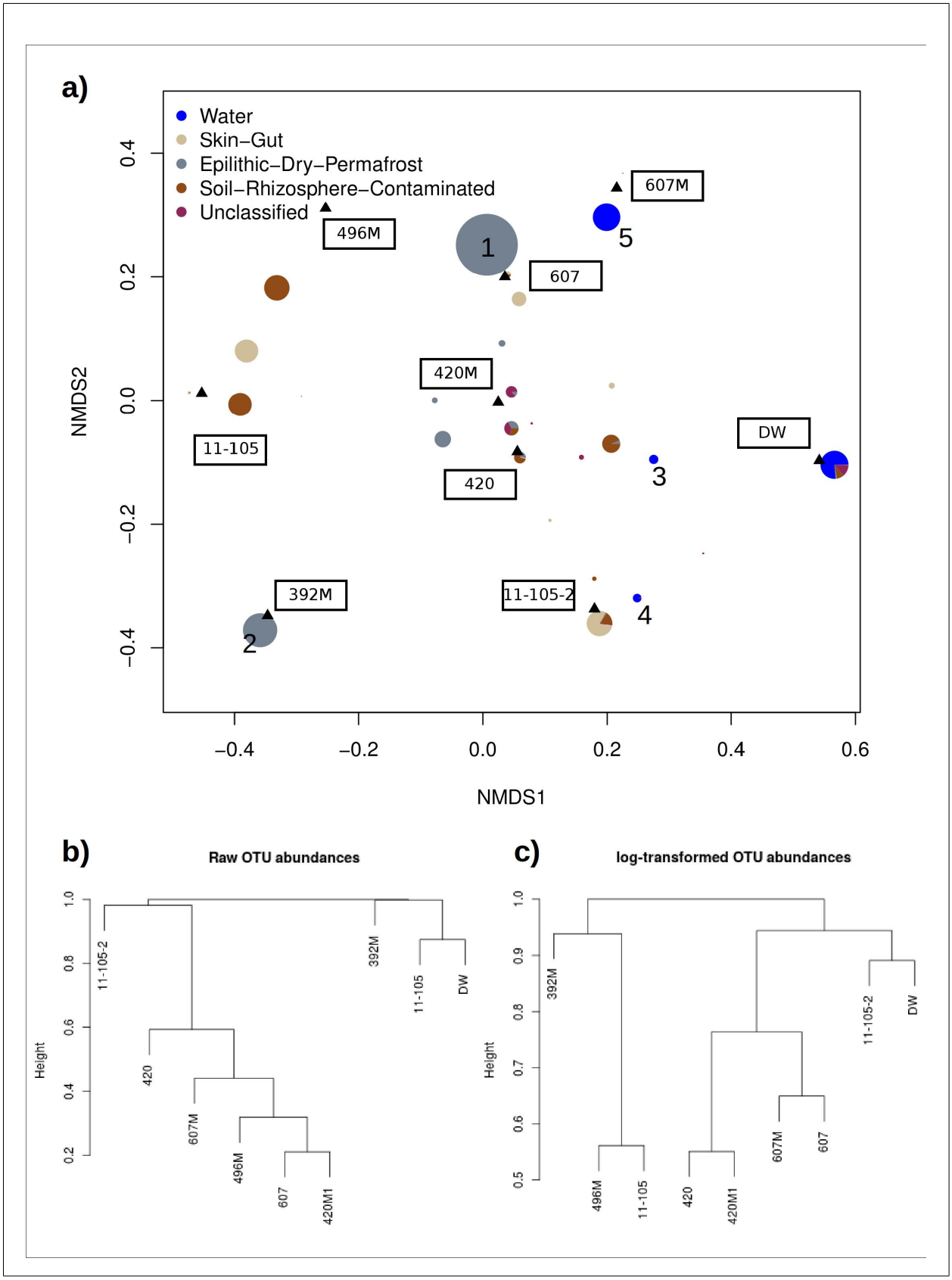

a)
Water
Skin–Gut
Epilithic–Dry–Permafrost
Soil–Rhizosphere–Contaminated
Unclassified
496M
607M
607
420M
11-105
420
DW
392M
11-105-2
1
2
3
4
5
NMDS1
NMDS2
b)
Raw OTU abundances
Height
c)
log-transformed OTU abundances
Height

Figure 7.5. Post-sequencing assessment of contaminants and MDA amplification biases
(**a**) Non-metric multidimensional scaling (NMDS) ordination showing the relationships between samples (triangles) and OTUs (circles). Circle areas represent relative abundances. Circle colour represents the OTU putative source, as described in **section 4.3.14**. In the case where several OTUs superimposed in the same region of the ordination (i.e. because they had a similar sample distribution) they were clustered together into a single circle whose radius corresponded to the sum of the original radii. For OTU clusters, the putative sources were displayed as a pie-chart, with the area for each colour corresponding to the proportion of OTUs in the cluster that shared each putative source. Numbers indicate OTUs that are further discussed in the text.
(**b**), (**c**) Hierarchical clustering of the samples before (**b**) and after (**c**) performing a logarithmic transformation of the OTU abundance data.

The drilling water sample (DW) was remarkably separated from the subsurface samples in the non-metric multidimensional scaling (NMDS) ordination (**Figure 7.5a**, x-axis). The most abundant OTUs in our dataset were similar to sequences isolated from epilithic, dry or permafrost habitats (**Figure 7.5a** – 1,2), whose environmental conditions share many similarities to those expected to be present in the BH10 borehole (see **sections 3.4.3, 1.1.3, 1.1.4**). Conversely, the OTUs belonging to the drilling water were similar to sequences isolated from water environments (**Figure 7.5a** – DW). Some water-related OTUs were shared between the drilling water and the deep subsurface samples (**Figure 7.5a** – 3,4), albeit they were far less abundant than the subsurface-related OTUs. This suggests that, even if a minimal degree of contamination due to the drilling water can be expected, it is in no case enough to mask the indigenous diversity of the subsurface samples. Human/animal microbiome related OTUs, which we have used as tracers for laboratory contaminations, were found in several samples, being predominant in one case (sample 11-105-2). For the rest of the samples, they were generally present in small proportions.

Hierarchical clustering confirmed the above results (**Figure 7.5b,c**). The MDA-amplified samples (samples 420M1, 607M) did not cluster together with their non-amplified counterparts (sample 420, 607) (**Figure 7.5b**), which shows that MDA indeed introduced enough quantitative biases to significantly alter the retrieved community composition. This detrimental side-effect was however mitigated after applying a logarithmic transformation to the OTU abundance data (Anderson *et al.*, 2006), a pre-processing step that is actually recommended for the multivariate analysis of ecological data (Anderson *et al.*, 2006; Kenkel, 2006). Logarithmic transformation not only resulted in the proper pairing of amplified and non-amplified samples, but also improved the overall clustering (**Figure 7.5c**).

7.3. Discussion

7.3.1. Tackling the signal-to-noise problem: a sensitive pipeline for deep subsurface metagenomics

There are several challenges that must be overcome in order to successfully perform metagenomic analyses on deep subsurface samples. The low biomass coupled to the adsorption of cells and nucleic acids to the mineral matrix will translate into very low DNA extraction yields, which in turn will increase the impact of even minute contaminations in the final results. Given the low extraction yields, an intermediate DNA amplification step will be also required, with the corresponding risk of further contamination and the introduction of taxonomic biases. Adequate measures must be taken to minimize all these risks, or at least to properly assess their actual impact in the retrieved results. Failing to do so would potentially lead to inaccurate – or even plainly erroneous – claims about the composition and function of the deep subsurface biosphere. We have successfully tested two DNA extraction protocols developed for deep subsurface samples, and adapted an UV decontamination step that ensures the sterility of the MDA DNA amplification process (**Figure 7.1**). We have also performed a test analysis of five deep subsurface samples in order to obtain an estimate of the degree of contamination that was remaining in our samples after subjecting them to our full analysis pipeline.

Even if only the innermost part of the core was sampled (see methods), drilling water (together with other contaminants from the drilling bit or from shallower sections of the borehole) could have still infiltrated through cracks during the drilling process. Since those contaminations are also coming from environmental sources (water, lubricants, shallower sections on the borehole), they can be especially difficult to identify, as their presence might appear “reasonable” or even “expected” to the researcher's eye. In order to assess the impact of this source of contamination in our study, a drilling water sample was analysed together with the subsurface samples.

We found little similarity between the drilling water library and the subsurface libraries (**Figure 7.5a**), which suggests that contamination during drilling is minimal. The presence of microbiome-related OTUs allowed us to estimate the degree of contamination introduced during sample manipulation and processing. Albeit microbiome-related OTUs were generally non-predominant, we decided to include a DNA extraction blank in the final analyses and subject it to MDA and sequencing (see next chapter), in order to confirm that sterility was maintained during the whole DNA processing pipeline.

Finally, our results showed that MDA introduced noticeable taxonomic biases (**Figure 7.5b**). However, the biases were greatly mitigated after applying a logarithmic transformation step (**Figure 7.5c**), which is a standard recommendation for most biostatistics pipelines. This solution minimizing the differences between the MDA-amplified samples and their non-amplified counterparts, while also rendering an overall better clustering. We therefore believe that, if appropriate measures are taken, the data originated from MDA-amplified samples is suitable for quantitative biostatistic analyses.

7.3.2. Poisson binomial filtering minimizes the impact of sequencing errors in microbial ecology studies

In this work we have presented and validated the Poisson binomial algorithm for filtering sequence reads based on their error probability distributions. We have also demonstrated that Poisson binomial filtering is especially useful in the context of gene-marker-based studies, such as the study of microbial populations by amplifying and sequencing their 16S rRNA gene.

We compared our algorithm with other five quality-filtering methods that are included as defaults in mainstream pipelines such as mothur, QIIME or USEARCH, by analysing mock and environmental datasets generated with three different sequencing platforms (**Figure 7.3**, **Annex 2 – Supplementary Note 5**). Our results show that, when coupled to a standard analysis pipeline that included chimera removal and clustering, PBF proved to be the most accurate algorithm for filtering marker-gene sequences. While retaining a large number of sequences, it also resulted in OTUs that were the closest to the true biological species present in the studied samples, and minimized the generation of spurious diversity and taxonomic biases.

Remarkably, this algorithm does not rely on any particular error model. Instead, it just derives the error probability distribution of a given sequence from the quality scores of its individual bases (**Figure 7.2**). To our knowledge, it is the first non-heuristic method to do so. The only assumption that our algorithm makes (which is shared with any other approach that utilizes quality scores) is that, for any given sequencing platform, the quality scores obtained during base calling will truly represent the probabilities of that base being wrong. This conceptual simplicity is one of its main advantages: as long as accurate quality scores are provided, Poisson binomial filtering will work in any present or future sequencing platform, with no need for further modifications.

In practice, quality-score calling ultimately depends on the sequencing platform manufacturer, and

its accuracy is also influenced by the choice of primers and library preparation methods (Loman *et al.*, 2012, Schirmer *et al.*, 2015). Nonetheless, we have shown that, for the three sequencing platforms studied in this work, Poisson binomial filtering was able to correctly discriminate between good and erroneous sequences based solely on quality score information.

The fact that our method only relies on quality scores means that it will only account for sequencing errors, but not other errors such as PCR substitutions. However, it has been described that sequencing errors are responsible for the majority of singletons generated in molecular ecology studies (Edgar, 2013, Tikhonov *et al.*, 2015). PCR chimeras are other source of spurious diversity, but dedicated algorithms such as UCHIME are able to accurately detect them.

During the course of this research, we have also focused on a source of taxonomic bias that may have affected the results of many molecular ecology studies. Most of the methods used for filtering and analysing marker-gene reads operate under the implicit (or even explicit, see Tikhonov *et al.*, 2015) assumption that the probability of having k errors is the same for all sequences, regardless of their origin. However, sequences from different taxa may have different length (for 454 and IonTorrent) and quality (for 454, IonTorrent and Illumina) distributions. This leads to the artificial enrichment of some taxa versus others during the quality filtering step (**Figure 7.4**), potentially compromising the quantitative interpretation of molecular ecology results obtained by high-throughput sequencing of marker-gene sequences. These biases are likely originated during base/quality calling: for instance, 454 reads show a systematic decrease in quality after homopolymer regions (Brockman *et al.*, 2008), which will penalize the taxa with longer homopolymer stretches on its 16S gene. We have nonetheless demonstrated that collapsing identical reads before the quality-filtering step greatly mitigates this issue (**Figure 7.4f**).

In summary, the methodologies presented in this work substantially improve the existing filtering approaches in terms of OTU accuracy, observed alpha diversity and observed community composition, delivering a more faithful representation of the original microbial communities present in the studied samples. Our algorithm is fast, easy to implement, and works for every sequencing platform constituting a valuable addition to all the existing pipelines for analysing microbial ecology data.

8. CHARACTERISATION OF THE BH10 SUBSURFACE ECOSYSTEM

8.1. Introduction

The borehole 10 (BH10) was drilled in 2012 during the IPBSL campaign. It was the deepest of the three boreholes studied during this work, reaching 613 metres deep, and its located farther from the Peña de Hierro mine pit than the BH8 borehole (see **Figure 3.1**). Probably due to this, the rocks in the BH10 are generally less fractured, resulting in a lower permeability. Localized fracture zones do however exist at the interfaces between the different stratigraphic units. Unlike the BH8, the BH10 aquifer hosts anoxic waters with a low redox potential. Stockwork pyrite is present, but it is unoxidised and in lower concentrations.

The BH10 is more recent than the BH8 borehole (see **section 3.2**), and thus it could be studied using state-of-the-art techniques. As concluded in **section 5.3.1**, a multi-technique approach is desirable in order to obtain the maximum possible information from deep subsurface core samples (**section 3.3**). In this chapter we focus on the analysis of high-throughput DNA sequencing data of both 16S amplicons and a full metagenome, obtained applying the protocols developed in **section 7**. Only samples below 100 mbsf were included, as the first 100 metres were drilled using destructive techniques which preclude the retrieval of intact cores. The information regarding the composition and potential metabolisms of the microbial communities found in the BH10 borehole was in turn integrated with the geochemical data obtained from the different samples in order to formulate hypotheses about the role, viability and survival mechanisms of the IPB deep subsurface microbiota.

The borehole BH10 was also studied by other microbial ecology techniques such as antibody microarrays, CARD-FISH or enrichment cultures. Those results were obtained by collaborators from the IPBSL project, and their analysis and discussion is therefore beyond the scope of this thesis. Nevertheless, a full multidisciplinar characterisation of the IPB subsurface ecosystem, which includes the results presented in this chapter, is currently underway (Amils *et al.*, in preparation).

8.2. Results

8.2.1. Geological, physicochemical and organic compound profiling of borehole BH10

Core samples were analysed by several techniques (see methods **sections 3.4.2, 3.4.3, 3.4.4**) in order to identify relevant substrates or products of different microbial metabolisms, as well as organic compounds that may be related to microbial presence or activity. This information was integrated with a stratigraphic profile of the BH10, including details on the presence of faults, shears and pyrite content throughout the borehole. The results are summarized in **Figure 8.1.**

Several faults were present between 100-150 mbsf. More definite fracture zones were found at 300, 420 and 500 mbsf. No faults were found below 500 mbsf, but rocks in that zone were more fractured than in the upper sections of the borehole. Bromide, which was added to the drilling water as a contamination tracer, remained at low concentrations except for a sample at 487 mbsf. Soluble iron was found at very low concentrations, except for a high peak between 250 and 350 mbsf, which was associated to an ammonium peak. Hydrogen was present in variable amounts, but significantly descended between 250 and 350 mbsf (being associated with the aforementioned increase in soluble iron and ammonium), between 400 and 450 mbsf (being associated with a fault) and below 500 mbsf (being associated with the more fractured Culm zone). An increase on methane concentrations was found between 250 and 350 mbsf, with another smaller increase being associated with the fault that marked the transition between the dark shale and the Culm zone at 500 mbsf. Sulphate, which is a known product of pyrite oxidation, was present at higher concentrations in the upper zone of the borehole, in which the presence of oxygen traces is more likely. The concentration of nitrate and nitrite generally descended with depth. Interestingly, nitrate and nitrite concentrations were negatively correlated between 400 and 500 mbsf.

Regarding organic compounds, formate and acetate, which are widespread electron donors for microbial metabolisms, were distributed in an irregular fashion, but followed a similar profile. Propionate was abundant at the shallower depths of the borehole, but was not detected below 300 mbsf. Oxalate was present at small concentrations, with peaks at 420 and 487 mbsf. Total organic carbon was generally low (1-2%), with a small peak at 420 mbsf, while sugar and proteins were distributed irregularly. Finally, palmitic and stearic acids, which can be related to extant/recent microbial presence (Sara Lincoln, personal communication), were detected at 353, 420 and 607 mbsf.

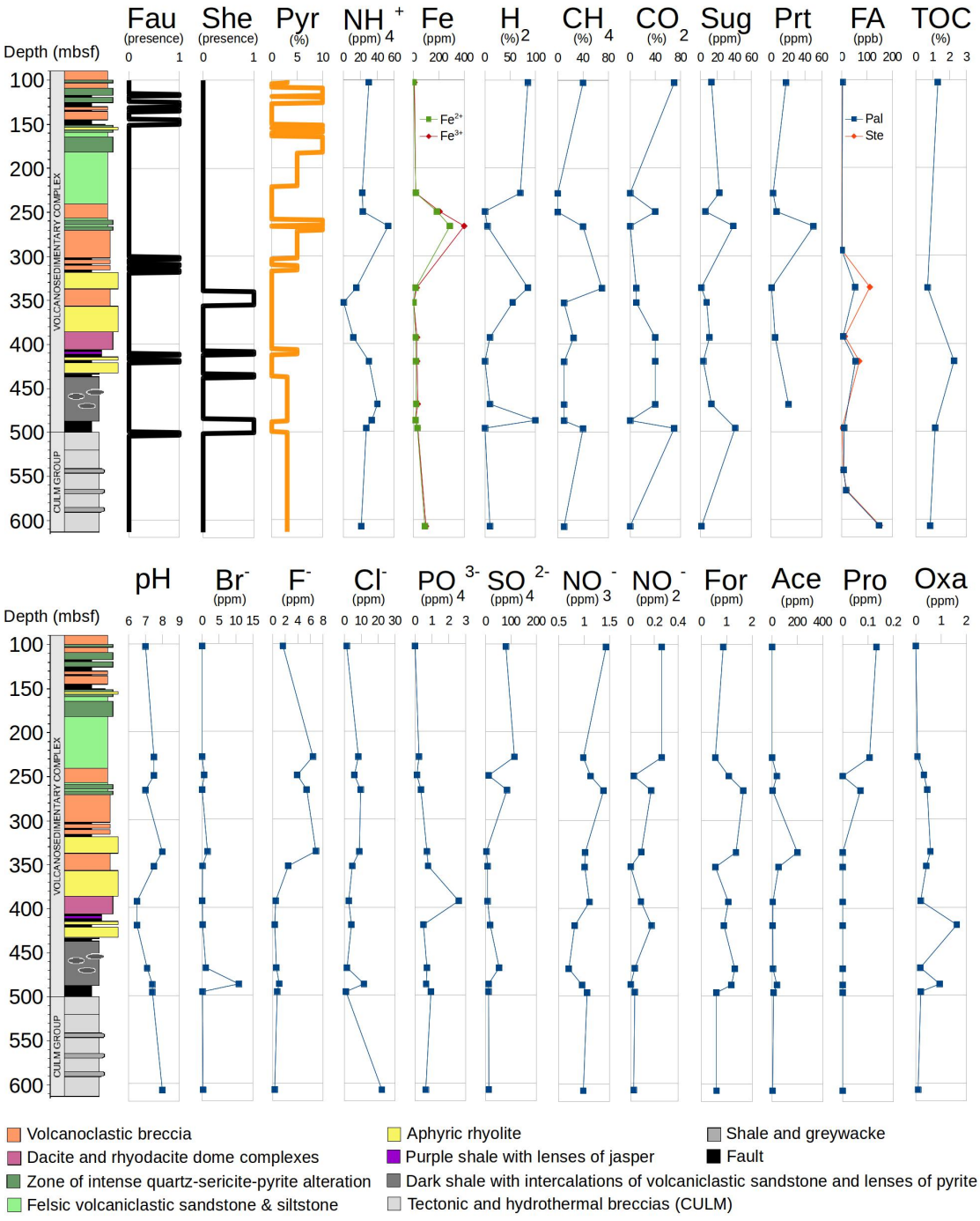


Figure 8.1: Geological, physicochemical and organic compound profiling of borehole BH10.

Abbreviations: Fau, faults; She, shears; Pye, pyrite content; Sug, total sugars; Prt, total proteins; FA, fatty acids; Pal, palmitic acid; Ste, stearic acid; TOC, Total organic carbon; For, formate; Ace, acetate; Pro, propionate; Oxa, oxalate.

8.2.2. Post-sequencing assessment of potential contamination sources and putative OTU origin environments

Two different controls were performed in order to assess if our samples had been contaminated during the drilling (DW control), DNA extraction, amplification or sequencing (IC control) processes. The incidence of contaminant OTUs on our Illumina MiSeq datasets was studied by hierarchical clustering of samples based on the presence/absence of 0.09 OTUs. This method is particularly stringent: the use of presence/absence scoring will link a sample to the contamination controls even if they only share low abundance OTUs, while the use of a 0.09 clustering distance for defining OTUs will ensure that this linkage is made beyond the genus level. The results presented in **Figure 8.2a** show that the DW control is completely separate from the rest of the samples, suggesting than drilling water was not a relevant source of contamination. Regarding the IC control, it clustered together with the samples from 228 and 353 mbsf, although both samples were far more similar between them than to the IC control. A certain degree of contamination might have therefore been introduced in those two samples during laboratory manipulation.

The putative OTU origin environments were studied as described in **section 7.2.4,** albeit with some differences (see **section 4.4.6**). As found in our preliminary analyses (**section 7.2.4**), the DW control clustered separately from the rest of the samples and the most abundant OTUs were similar to sequences obtained from Epilithic-Dry-Permafrost environments (**Figure 8.2b**). OTUs related to sludge, wastewater, contaminated, soil or rhizosphere environments also dominated in some samples. There was a small presence of OTUs related to acid mine drainage or metal-contaminated environments. OTUs related to human/animal microbiome were also detected. Finally, OTUs related to water environments were particularly dominant in the DW control, albeit they were also present in small proportions in some borehole samples. As discussed in the previous chapter (**section 7.3.1**), this could indicate a limited contamination during the drilling process.

8.2.3. Phylotype-level profile of borehole BH10

The phylotype-level profile of borehole BH10 is shown in **Figure 8.2c** The most abundant phylotypes were *Actinobacteria*, *Alphaproteobacteria* and *Cyanobacteria*. Actinobacteria were more abundant in the shallower samples, and cyanobacteria were more abundant in the deeper samples; apart from this, no other clear trend could be visually identified. Regarding the two controls, both were distinguishable from the rest of the samples: the DW control was by far the most diverse, while the IC control was less diverse and dominated by gammaproteobacteria.

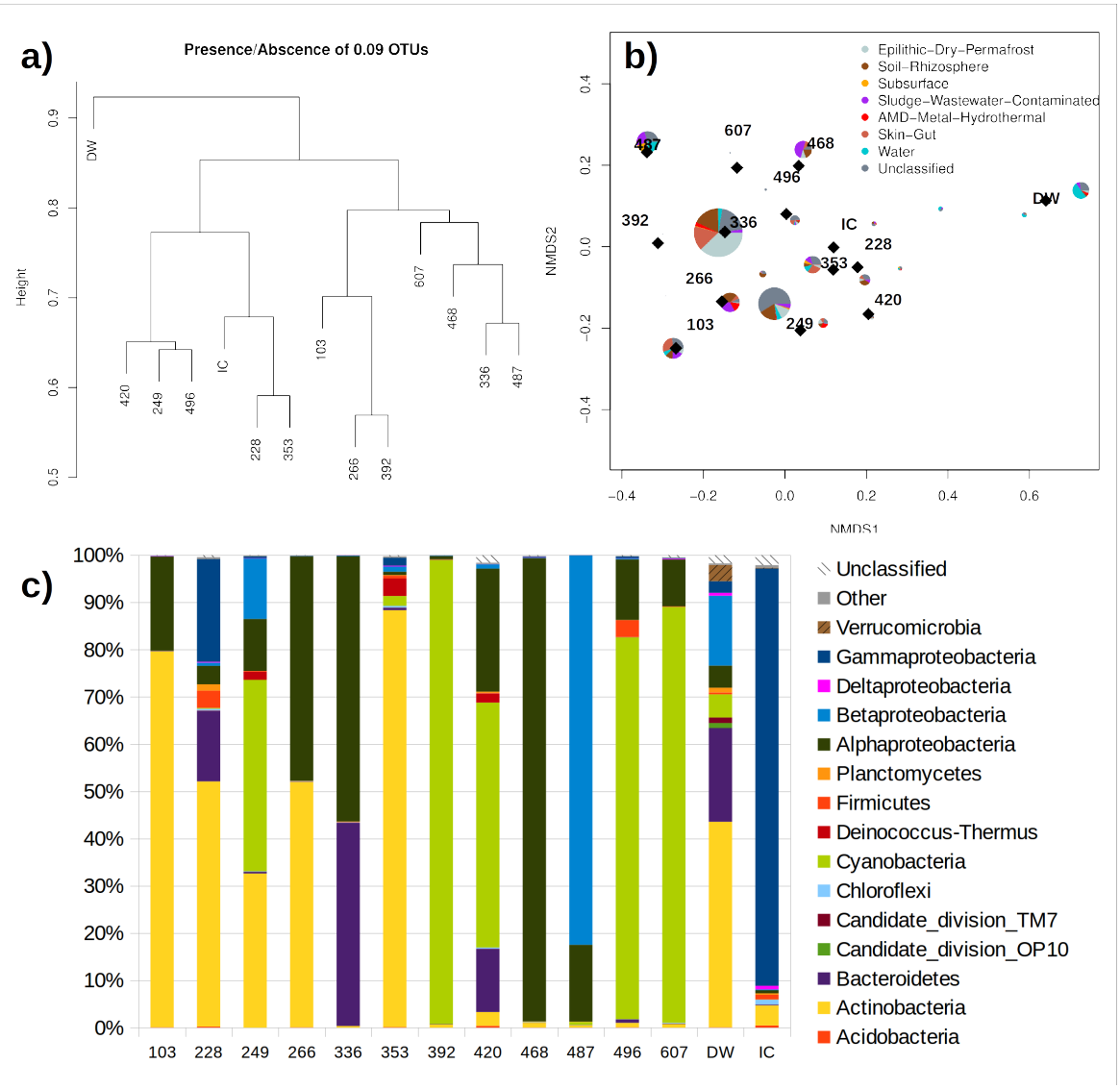


Figure 8.2: 16S rRNA gene profiling of borehole BH10.

(a) Hierarchical clustering of BH10 samples, using presence/absence of OTUs defined at a 9% distance.
(b) Non-metric multidimensional scaling (NMDS) ordination showing the relationships between samples (diamonds) and OTUs (circles). Circle areas represent relative abundances. For each OTU, circle colours represent the proportion of blast hits assigned to the different putative sources, as described in **section 4.4.6** (Note that putative sources were assigned in a different way than in **Figure 7.5**). In the case where several OTUs superimposed in the same region of the ordination (i.e. because they had a similar sample distribution) they were clustered together into a single circle whose radius corresponded to the sum of the original radii, and whose colour composition corresponded to the proportion of blast hits from each constituent OTU that were assigned to each putative source, weighed by the abundance of said OTU.
(c) Taxonomic composition at the phylum level (with *Proteobacteria* expanded into classes) of BH10 samples. Sample names indicate sampling depth, except in DW (drilling water control) and IC (internal laboratory control).

Table 8.1. Nearest environmental clone, isolate and type strain of the 50 most abundant OTUs from the BH10 borehole. OTUs appearing in the IC or DW controls were excluded. In the cases when the closest isolate was similar to the closest type strain, it was omitted from the table.

Nearest neighbor	Score	Closest Isolate	Score	Closest Type strain	Score	Putative isolation sources
Uncultured bacterium; YF829	0.876	Calothrix sp. ANTLPR2.4	0.817	Crinalium epipsammum (T); SAG 22.89	0.701	Epilithic-Dry-Permafrost; Soil-Rhizosphere; Skin-Gut
Uncultured cyanobacterium; B1_7_1B_78	0.971	Toxopsis calypsius PLF.1	0.929	Crinalium epipsammum (T); SAG 22.89	0.742	Epilithic-Dry-Permafrost; Soil-Rhizosphere; Skin-Gut
Denitrifying Fe-oxidizing bacteria; BrG1	1.000	---		Acidovorax defluvii (T); BSB411	1.000	Sludge-Wastewater-Contaminated; Unclassified; Subsurface; Water
Arthrobacter oryzae; S3B	1.000	---		Arthrobacter oryzae (T); KV-651	0.975	Unclassified; Soil-Rhizosphere; Epilithic-Dry-Permafrost; Sludge-Wastewater-Contaminated; Water
Tessaracoccus sp. R-36529	0.933	---		Tessaracoccus flavescens (T); type strain:SST-39	0.833	Skin-Gut; Sludge-Wastewater-Contaminated; Unclassified; Soil-Rhizosphere; Epilithic-Dry-Permafrost
Uncultured bacterium; Y2903	0.950	Leptolyngbya sp. CAWBG532	0.837	Planktothricoides raciborskii (T); NIES-207	0.754	Epilithic-Dry-Permafrost; Soil-Rhizosphere; Skin-Gut; Unclassified; AMD-Metal-Hydrothermal
Uncultured bacterium; MI224	1.000	---		Bradyrhizobium elkanii (T); USDA 76	1.000	Soil-Rhizosphere; AMD-Metal-Hydrothermal; Sludge-Wastewater-Contaminated; Unclassified
Uncultured bacterium; YL1022	0.971	Chroococcidiopsis sp. CCMEE 246	0.859	Planktothricoides raciborskii (T); NIES-207	0.710	Skin-Gut; Epilithic-Dry-Permafrost; Soil-Rhizosphere; Subsurface; Unclassified
Uncultured bacterium; HDB_SIPA1453	1.000	Sediminibacterium sp. I-32	1.000	Chitinophaga niastensis (T); JS16-4	0.813	Unclassified; Skin-Gut
Uncultured bacterium; ftbw109a10c1	1.000	---		Rothia mucilaginosa (T); DSM	1.000	Skin-Gut; Subsurface; Water
Sphingomonas mucosissima; XFB-BR	1.000	---		Sphingomonas dokdonensis (T); DS-4	0.929	Unclassified; Water; Skin-Gut; Epilithic-Dry-Permafrost; Soil-Rhizosphere
Uncultured bacterium; BE325FW032701CFS_hde1-3	1.000	---		Caulobacter vibrioides (T); CB51	1.000	Unclassified; Sludge-Wastewater-Contaminated; Skin-Gut; AMD-Metal-Hydrothermal; Epilithic-Dry-Permafrost; Soil-Rhizosphere; Water
Alpha proteobacterium OS-09	1.000	---		Caulobacter leidyi (T); ATCC 15260(T)	1.000	Soil-Rhizosphere; Water; Unclassified; Epilithic-Dry-Permafrost
Phormidium autumnale CCALA 143	1.000	---		Crinalium epipsammum (T); SAG 22.89	0.813	Epilithic-Dry-Permafrost; Soil-Rhizosphere; Unclassified
Uncultured bacterium; ftbw207c03c1	0.970	Haemophilus parainfluenzae; HK 2149	0.949	Haemophilus parainfluenzae (T); CCUG 12836	0.915	Skin-Gut; Unclassified
Uncultured bacterium; rcd2068c11c1	0.874	Roseomonas sp. S10(2011)	0.849	Roseomonas pecuniae (T); N75	0.824	Unclassified; Epilithic-Dry-Permafrost; Skin-Gut; Subsurface
Hymenobacter flocculans A2-50A	1.000	---		Hymenobacter ocellatus (T); type strain: Myx 2105	0.816	Unclassified; Water; AMD-Metal-Hydrothermal; Skin-Gut; Soil-Rhizosphere
Methylbium sp. W125	0.950	---		Rhizobacter dauci (T); H6 (=ATCC43778)	0.925	Epilithic-Dry-Permafrost; Soil-Rhizosphere; Unclassified; Skin-Gut
Alpha proteobacterium OR-84	1.000	---		Inquilinus ginsengisoli (T); Gsoil 080	0.737	Unclassified; Water; Sludge-Wastewater-Contaminated; Subsurface; AMD-Metal-Hydrothermal; Soil-Rhizosphere
Uncultured bacterium; 003	0.942	Anabaena torulosa BF1	0.855	Crinalium epipsammum (T); SAG 22.89	0.718	Epilithic-Dry-Permafrost; Unclassified
Allihoeffea aestuarii; 2WW	1.000	---		Hoeflea alexandrii (T); type strain:AM1V30	0.809	Unclassified; Epilithic-Dry-Permafrost; Soil-Rhizosphere; AMD-Metal-Hydrothermal
Dermacoccus profundi (T); MT2.2	1.000	---		Dermacoccus barathri (T); MT2.1	1.000	Unclassified; Skin-Gut; Soil-Rhizosphere; Epilithic-Dry-Permafrost; Sludge-Wastewater-Contaminated
Uncultured bacterium; 155	0.924	Phenylobacterium sp. C16-Sin106	0.907	Phenylobacterium falsum (T); type strain:AC-49	0.886	Water; Sludge-Wastewater-Contaminated; Epilithic-Dry-Permafrost; Subsurface; Soil-Rhizosphere; AMD-Metal-Hydrothermal; Unclassified
Acidiphilium cryptum; Kuesel IFV	1.000	---		Acidiphilium multivorum (T); AIU 301	1.000	AMD-Metal-Hydrothermal; Unclassified
Uncultured Deinococcales bacterium; A2_1a_2B_6	1.000	---		Truepera radiovictrix (T); RQ-24	0.803	Epilithic-Dry-Permafrost; Skin-Gut; Unclassified
Streptococcus oralis (T); ATCC 35037	1.000	---		Streptococcus oralis (T); ATCC 35037	1.000	Skin-Gut; Unclassified; Water
Uncultured bacterium; rbh12b06;	1.000	Corynebacterineae bacterium NML 120705	0.896	Corynebacterium aquilae (T); type strain: CECT 5993	0.762	Skin-Gut; Unclassified
Arthrobacter sulfureus (T); DSM 20167	1.000	---		Arthrobacter sulfureus (T); DSM 20167	1.000	Unclassified; Epilithic-Dry-Permafrost; Water; Soil-Rhizosphere
Uncultured bacterium; JSC9-B11	1.000	---		Rubrobacter radiotolerans; type strain:DSM 5868	0.766	Epilithic-Dry-Permafrost; Skin-Gut; Soil-Rhizosphere; Unclassified
Uncultured bacterium; rcd512a10c1	0.915	Hymenobacter sp. R2-4	0.843	Hymenobacter norwichensis (T); type strain: NS60	0.804	Epilithic-Dry-Permafrost; Skin-Gut; Unclassified; Soil-Rhizosphere; Water
Uncultured bacterium; VB35	1.000	Microcoleus sp. PCC 7113	0.878	Crinalium epipsammum (T); SAG 22.89	0.727	Epilithic-Dry-Permafrost; Soil-Rhizosphere
Acinetobacter junii; BB1A	1.000	---		Acinetobacter junii (T); DSM 6964	1.000	Unclassified; Skin-Gut; Water; Sludge-Wastewater-Contaminated; Soil-Rhizosphere; Subsurface
Uncultured bacterium; HDB_SIPP670	0.983	Flavisolibacter sp. MDT2-37	0.915	Flavisolibacter ginsengiterrae (T); Gsoil 492	0.889	Skin-Gut; Epilithic-Dry-Permafrost; Sludge-Wastewater-Contaminated; Soil-Rhizosphere; Subsurface; Unclassified
Spirosoma linguale; ATCC 23276	0.948	---		Spirosoma spitsbergense (T); SPM-9	0.922	Unclassified; Epilithic-Dry-Permafrost; Sludge-Wastewater-Contaminated; Skin-Gut; Soil-Rhizosphere
Uncultured bacterium; V60_14a1	0.979	Chroococcidiopsis sp. BB79.2; SAG 2023	0.979	Crinalium epipsammum (T); SAG 22.89	0.742	Epilithic-Dry-Permafrost; Skin-Gut

8.2.4. Individual OTU analysis

While the inter-sample comparisons shown in **section 8.2.2** were performed using OTU clustering distances of 0.09 in order to favour the clustering of somewhat dissimilar samples, the results presented in this section were obtained using 0.03 (species level) OTU clustering distances. OTUs appearing in the IC or DW controls were excluded from this analysis. The inferred evolutionary relationships between borehole BH10 OTUs and their closest phylogenetic relatives is presented in **Annex 4 – Figure S1**. More detailed information on the most abundant OTUs can be found in **Table 8.1** and **Annex 3 - Table S3**.

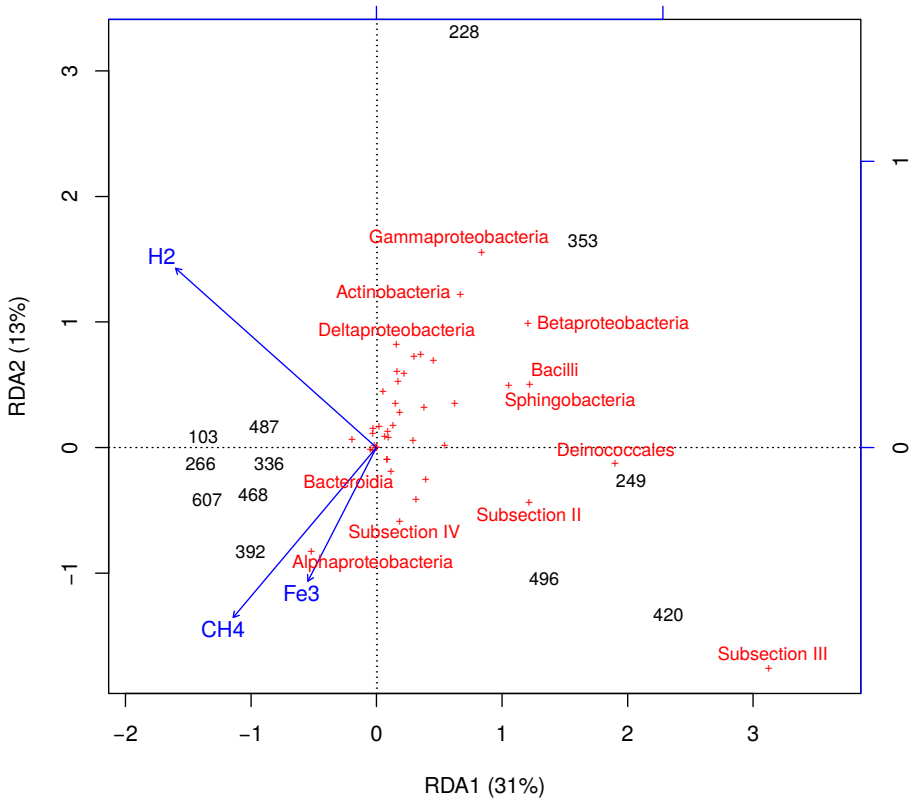
The most abundant OTUs in borehole BH10 belonged to the *Cyanobacteria* phylum (**Table 8.1**), and were related to known endolithic and hypolithic cyanobacterial genera such as *Calothrix* (Pentecost, 2005), the filamentous *Crinalium* (Stal, 2007), *Chroococcidiopsis* (Baqué *et al*, 2013) or the cave-dwellers *Toxopsis* (Lamprinou *et al.* 2012) and *Leptolyngbia* (Bruno *et al.*,

2009). Several non-cyanobacterial OTUs were also present in high abundances, including one related to an iron-oxidizing, nitrate-reducing *Acidovorax* (Straub *et al.*, 1996), and the ubiquitous soil and subsurface bacteria *Arthrobacter* (Crocker *et al.*, 2000; van Waasbergen et al 2000; Kageyama *et al.*, 2008), *Bradyrhizobium* (Wang *et al.*, 2013) and *Sediminibacterium* (Bollmann *et al.*, 2010). Interestingly, an OTU related to the *Tessaracoccus* genus (see **section 6.3.4**) was also found among the most abundant OTUs.

8.2.5. Effect of environmental variables on the microbial composition of the BH10 borehole

Several combinations of explanatory environmental variables were tested for fitting a redundancy analysis model together with the phylotype abundance data. Finally, hydrogen (p=0.002), ferric iron (p=0.051) and methane (p=0.054) were chosen as explanatory variables, being able to explain 43% of the variation in phylotype abundances across samples (**Figure 8.3**).

Figure 8.3. Effect of environmental variables on the microbial community composition in the BH10 borehole.



Redundancy analysis (RDA) showing the variation of microbial community composition with respect to three environmental variables. The three environmental constraints used to build our model were found to be significant by a permutation test. Black numbers, sample names; red crosses, phylotypes (class level); blue arrows, environmental parameters. Phylotype names are only included for phylotypes with non-uniform distributions.

Methane and soluble ferric iron concentrations were correlated, while both were independent from hydrogen concentrations (**Figure 8.3**). There was a strong negative correlation between hydrogen concentrations and the relative abundance of members of the Subsection III (*Leptolyngbia*, *Crinalium*) of the *Cyanobacteria* phylum, and to a lesser extent the Subsection II (*Chroococcidiopsis*) and the *Deinococcales* order. On the other hand, the relative abundance of gammaproteobacteria (*Haemophilus*, *Acinetobacter*, *Pseudomonas*), actinobacteria (*Tessaracoccus*, *Arthrobacter*, *Methylibium*), betaproteobacteria *(Acidovorax*) and deltaproteobaceria (*Desulfosoma*) was negatively correlated with ferric iron and methane concentrations. Finally, the relative abundance of alphaproteobacteria (*Bradirhizobium*, *Sphingomonas*, *Caulobacter*, among others) was weakly correlated with ferric iron and methane concentrations. Let the reader bear in mind that the genera affiliations presented in this paragraph are only approximate (see **Table 8.1**), and that the primers used for 16S rRNA gene amplification, albeit theoretically able to amplify the archaeal 16S rRNA gene, favour the amplification of bacterial species (Alejandro Arce-Rodríguez, personal communication, see also Bohorquez *et al.*, 2012).

8.2.6. Metagenomic analysis of sample 420

Due to the strong negative correlation found between the Subsection III of the *Cyanobacteria* phylum and hydrogen concentrations, the sample retrieved at 420 mbsf (sample 420) was chosen for further metagenomic analysis. The taxonomy of the metagenomic reads was consistent with the one found for sample 420 during the 16S analysis, with a predominance of cyanobacteria, alphaproteobacteria and bacteroidetes (**Figure 8.4a**). However, 23.50% of the reads had an eukaryotic origin, most of which were further classified as ascomycotal. The presence of archaeal or viral reads was negligible (**Figure 8.4a**). A more detailed taxonomic classification of the four most abundant taxa found in sample 420 is given in **Figure 8.4c**. According to MG-RAST classification, the cyanobacterial population was mainly composed of reads from the *Nostocales* and *Chroococcales* orders; the ascomycotal population was mainly composed of reads from *Eurotiomycetes*, *Sordariomycetes* and *Dothideomycetes* classes; the alphaproteobacterial population was mainly composed of reads from the *Rhizobiales* order, and bacteroidetes were mainly composed of reads from the *Cytophagales* and *Sphingobacteriales* orders. The results regarding order-level bacterial taxonomy were again consistent to the ones found by 16S rRNA gene amplicon sequencing of sample 420 (**Table 8.1, Annex 3 - Table S3**), except for the cyanobacterial population, for which the most abundant 16S OTU was a member of the *Leptolyngbia* (order

Oscillatoriales) genus. A broad overview of the relative abundance of different metabolic subsystems in the metagenome of sample 420 is given in **Figure 8.4b.** After manually inspecting all the annotations, many specific functions were found to be of interest for discussing the potential origin and survival strategies of the microbial community present in sample 420 (Table 8.2).

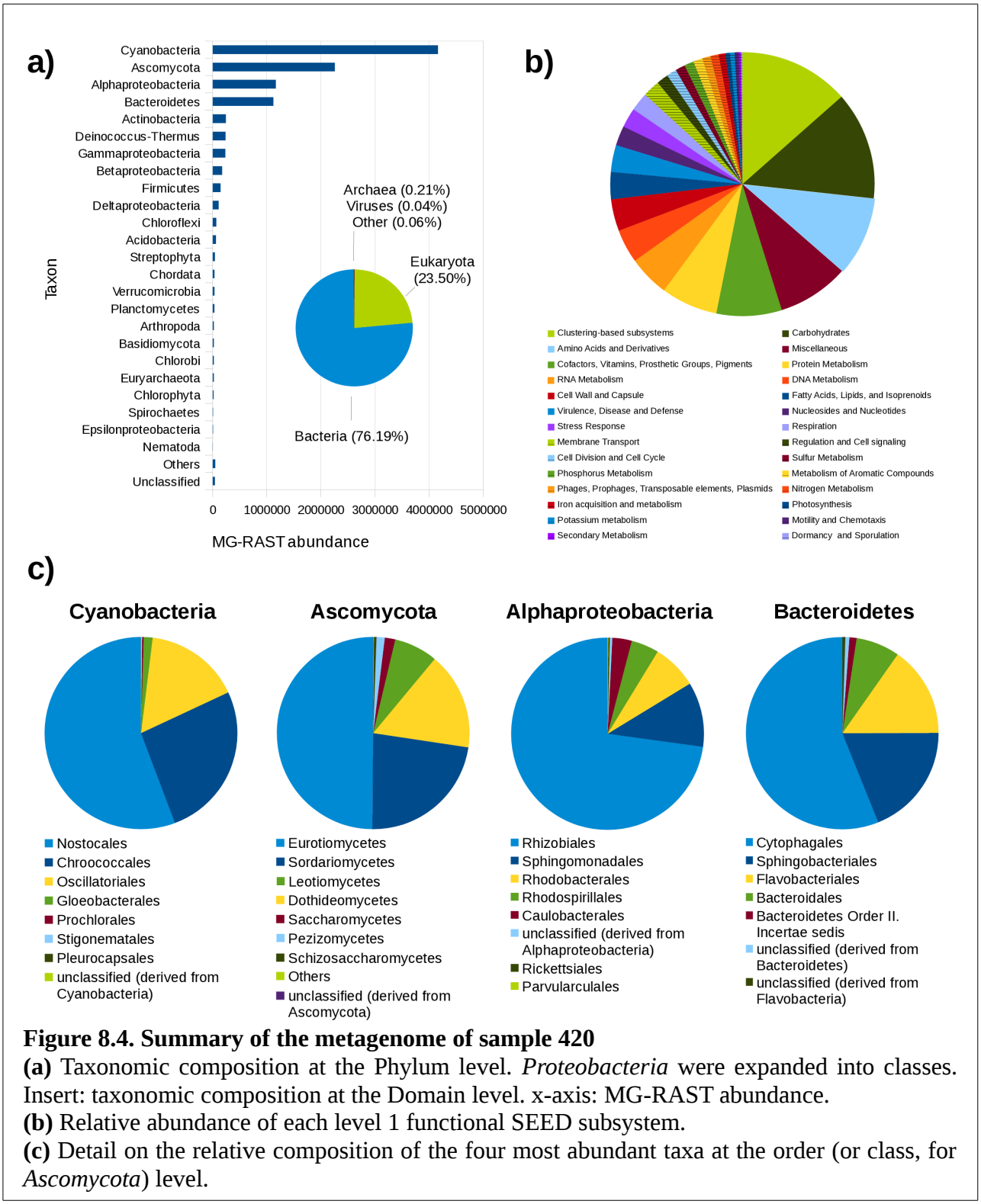


Figure 8.4. Summary of the metagenome of sample 420
(a) Taxonomic composition at the Phylum level. *Proteobacteria* were expanded into classes. Insert: taxonomic composition at the Domain level. x-axis: MG-RAST abundance.
(b) Relative abundance of each level 1 functional SEED subsystem.
(c) Detail on the relative composition of the four most abundant taxa at the order (or class, for *Ascomycota*) level.

Table 8.2: Abundance of key metabolic functions in the four most abundant taxa found in the metagenome of sample 420. Values: MG-RAST abundances.

Transporters	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
2-keto-3-deoxygluconate permease (KDG permease)	0	0	1	0	1
ABC-type Fe3+-siderophore transport system, permease 2 component	234	0	3	0	30
ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	8	0	96	0	73
ABC-type sugar transport system, permease component	307	0	21	0	92
ABC-type transport system involved in resistance to organic solvents, permease component USSDB6A	0	0	133	0	24
ABC-type tungstate transport system, permease protein	0	0	35	0	14
Alkanesulfonates transport system permease protein	99	0	726	0	168
Alpha-glucoside transport system permease protein AglG	8	0	15	0	17
Amino acid permease in hypothetical Actinobacterial gene cluster	20	0	0	0	22
ammonium/methylammonium permease	82	0	0	0	15
Branched-chain amino acid transport system permease protein LivM (TC 3.A.1.4.1)	1725	0	890	0	568
Cytosine/purine/uracil/thiamine/allantoin permease family protein	0	156	196	14	294
D-beta-hydroxybutyrate permease	0	0	76	0	23
D-galactarate permease	0	0	185	0	3
D-glucarate permease	0	0	63	0	10
Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	0	0	674	103	614
Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	0	0	401	19	516
Ethanolamine permease	0	0	54	0	138
Fucose permease	0	0	18	134	30
Gluconate permease	0	0	0	68	99
Glutamate Aspartate transport system permease protein GltK (TC 3.A.1.3.4)	0	0	94	0	17
Glycine betaine ABC transport system, permease protein OpuAB	0	0	2	0	57
Glycolate permease	0	0	83	0	49
High-affinity branched-chain amino acid transport system permease protein LivH (TC 3.A.1.4.1)	1178	0	1653	0	813
Homolog of fucose/glucose/galactose permeases	0	0	31	3	15
Inositol transport system permease protein	89	0	15	108	64
Iron(III) dicitrate transport system permease protein FecD (TC 3.A.1.14.1)	354	0	5	2	235
L-lactate permease	247	0	136	124	232
L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)	1	0	97	3	476
Monomethylamine permease	0	0	14	0	0
Multiple sugar ABC transporter, membrane-spanning permease protein MsmF	74	0	86	0	90
N-acetyl glucosamine transporter, NagP	0	0	3	572	79
Na+/H+-dicarboxylate symporters	0	0	85	0	57
Niacin transporter NiaP	0	0	151	0	167
Nickel transport system permease protein NikC (TC 3.A.1.5.3)	0	0	15	0	26
Nitrate ABC transporter, permease protein	1025	0	136	0	24
Nitrate/nitrite transporter	1034	49	200	113	220
Nucleoside permease NupC	96	0	50	0	78
Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)	1761	0	201	86	211
Phenylalanine-specific permease	0	0	1	0	9
Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	1666	0	244	25	365
Phosphonate ABC transporter permease protein phnE (TC 3.A.1.9.1)	229	0	43	0	169
Predicted cobalt transporter CbtA	0	0	14	0	7
Predicted erythritol ABC transporter 1, permease component 1	0	0	81	0	0
Predicted nucleoside ABC transporter, permease 1 component	0	0	12	0	3
Predicted sialic acid transporter	0	0	0	51	21
Predicted sodium-dependent mannose transporter	0	0	0	176	53
Predicted thiamin transporter PnuT	0	0	0	1	0
Predicted trehalose permease, MFS family	0	0	0	11	25
Probable 3-phenylpropionic acid transporter	0	0	44	0	1
Probable low-affinity inorganic phosphate transporter	1334	177	127	97	29
Putative deoxyribose-specific ABC transporter, permease protein	0	0	65	0	113
Putative glutathione transporter, permease component	0	0	631	0	113
Quinate permease	0	47	0	0	0
Sialic acid transporter (permease) NanT	0	0	88	0	32
Sodium-dependent phosphate transporter	2	0	126	4	28
Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)	216	0	69	0	133
sugar ABC transporter, permease protein	1672	0	1022	8	583
Sulfate permease	998	0	152	23	335
Sulfate transporter family protein in cluster with carbonic anhydrase	0	0	16	0	79
Taurine transport system permease protein TauC	0	0	9	0	67
Taurine transporter substrate-binding protein	0	0	2	0	3
Urea ABC transporter, permease protein UrtB	262	0	631	7	84
Various polyols ABC transporter, permease component 1	0	0	245	0	159
Xanthine/uracil/thiamine/ascorbate permease family protein	437	0	14	67	422
Xanthosine permease	0	0	0	120	56
Xylose ABC transporter, permease protein XylH	0	0	371	0	53

Catabolism of sugars, lipids and aminoacids	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Xylose utilization	553	46	2118	1983	2384
Valine degradation	607	1232	7330	4835	3372
Tryptophan catabolism	181	304	387	556	558
Trehalose Uptake and Utilization	4036	188	97	705	648
TCA Cycle	16654	2966	5862	5939	7967
Sucrose utilization	680	10	200	162	441
Pentose phosphate pathway	17943	1251	4214	2419	6030
Methionine Degradation	5060	507	2152	2456	2364
Melibiose Utilization	0	0	51	207	491
Mannitol Utilization	680	31	635	9	464
Maltose and Maltodextrin Utilization	17452	91	3644	3142	5971
Mannose Metabolism	8968	886	929	2664	1562
Leucine Degradation and HMG-CoA Metabolism	3310	1513	3578	5419	3759
Lactate utilization	247	0	495	1107	929
L-rhamnose utilization	14	13	885	4574	1946
L-fucose utilization	28	76	307	739	592
L-Arabinose utilization	0	64	618	1865	1923
Isoleucine degradation	681	1058	8852	7435	4550
Inositol catabolism	274	4	418	1090	2716
Glycolysis and Gluconeogenesis	24197	1741	4152	3785	5634
Glycogen metabolism	16402	444	2325	1831	2824
Fatty acid degradation regulons	3902	728	9485	1765	4184
Entner-Doudoroff Pathway	17845	1414	3771	4856	4546
D-Tagatose and Galactitol Utilization	1207	30	74	475	328
D-Sorbitol(D-Glucitol) and L-Sorbose Utilization	19	0	31	0	127
D-ribose utilization	2831	52	1137	635	1913
D-gluconate and ketogluconates metabolism	2694	252	841	2451	1678
D-Galacturonate and D-Glucuronate Utilization	3319	47	1167	3682	2040
D-galactonate catabolism	0	30	355	260	399
D-galactarate, D-glucarate and D-glycerate catabolism	678	43	1278	130	958
D-allose utilization	0	0	2	0	4
Chitin and N-acetylglucosamine utilization	3637	149	379	3458	2099
Branched chain amino acid degradation regulons	1313	1496	5878	6043	4210
Beta-glucuronide utilization	0	0	0	4	4
Beta-Glucoside Metabolism	0	0	174	394	794
Aromatic amino acid degradation	801	66	1697	800	953
Alginate metabolism	5188	374	989	1853	1246

Table 8.2: Abundance of key metabolic functions in the four most abundant taxa found in the metagenome of sample 420. Values: MG-RAST abundances.

Fermentation	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Fermentations: Lactate	5438	330	2020	374	3654
Fermentations: Mixed acid	5005	149	1712	1291	2793
Butyrate metabolism cluster	47	221	5097	2158	2102
Acetoin, butanediol metabolism	2669	314	438	524	1069
Acetone Butanol Ethanol Synthesis	462	484	4559	2481	3778
Acetyl-CoA fermentation to Butyrate	112	492	9467	3517	3691

Electron donating reactions	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Na(+)-translocating NADH-quinone oxidoreductase and rnf-like group of electron transport complexes	0	0	1	47	66
Hydrogenases	713	0	494	21	220
Sulfur oxidation	1555	0	1205	100	576
Formate hydrogenase	2461	111	2856	481	1492
Methanol dehydrogenase large subunit protein (EC 1.1.99.8)	0	0	404	0	66
Glucose-methanol-choline (GMC) oxidoreductase:NAD binding site	462	0	410	2126	368
Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	21	0	253	28	208
D-Lactate dehydrogenase (EC 1.1.2.5)	0	0	8	0	16
D-amino acid dehydrogenase small subunit (EC 1.4.99.1)	553	0	395	2	150
Glutamate dehydrogenases	1901	410	401	705	1162
Glucose dehydrogenase, PQQ-dependent (EC 1.1.5.2)	0	0	320	968	150
Glycerol dehydrogenase (EC 1.1.1.6)	1131	0	5	0	101
Glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	0	157	58	127	165
L-lactate dehydrogenase (EC 1.1.2.3)	0	0	291	0	97
NADH dehydrogenase (EC 1.6.99.3)	3065	0	264	551	405
Proline dehydrogenase (EC 1.5.99.8) (Proline oxidase)	3176	0	353	0	145
Respiratory Complex I	17512	1007	5299	5382	5352
Dehydrogenase complexes	4660	1151	4189	5676	3968
CO Dehydrogenase	0	0	2160	5	619

Electron accepting reactions	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Trimethylamine-N-oxide reductase (Cytochrome c) (EC 1.7.2.3)	0	0	9	0	106
Terminal cytochrome O ubiquinol oxidase	0	0	726	0	214
Terminal cytochrome d ubiquinol oxidases	456	0	1144	152	710
Terminal cytochrome C oxidases	4105	349	2382	1264	1516
Respiratory nitrate reductase alpha/beta/gamma chains (EC 1.7.99.4)	0	0	6	0	51
Nitrous-oxide reductase (EC 1.7.99.6)	0	0	0	2	0
Nitric-oxide reductase (EC 1.7.99.7), quinol-dependent	1	0	1	31	34
Cytochrome c551 peroxidase (EC 1.11.1.5)	0	0	11	43	166
Cytochrome c-type protein NapC	0	0	3	0	6
Copper-containing nitrite reductase (EC 1.7.2.1)	0	0	1	30	2
Anaerobic DMSO reductase	0	0	69	0	153
Anaerobic dehydrogenases, typically selenocysteine-containing	0	0	254	0	162

Related to Reactive Oxygen Species	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)	0	0	18	1	20
Superoxide dismutase [Fe] (EC 1.15.1.1)	136	0	113	0	95
Superoxide dismutase [Mn] (EC 1.15.1.1)	2	14	1	0	49
Similar to non-heme chloroperoxidase	615	0	8	0	82
Similar to non-heme chloroperoxidase, sll5080 homolog	54	0	8	0	82
Predicted iron-dependent peroxidase, Dyp-type family	0	0	17	0	72
Peroxidase (EC 1.11.1.7)	38	28	327	226	707
Multicopper oxidase	733	232	905	524	519
Manganese superoxide dismutase (EC 1.15.1.1)	625	0	120	321	779
Cytochrome c551 peroxidase (EC 1.11.1.5)	0	0	11	43	166
Catalase (EC 1.11.1.6)	218	34	517	532	3267

Resistance to environmental stresses	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Resistance to chromium compounds	0	0	299	30	523
Synthesis of osmoregulated periplasmic glucans	0	0	1183	18	108
Copper homeostasis: copper tolerance	133	0	347	43	265
Choline and Betaine Uptake and Betaine Biosynthesis	295	199	1056	35	3981
Cadmium resistance	611	0	0	1	563
Cobalt-zinc-cadmium resistance	39391	0	14368	13570	11524
Protection from Reactive Oxygen Species	1017	62	1106	1123	5034
Mercury resistance operon	1850	0	165	12	214
Acid resistance mechanisms	2752	233	394	245	341
Mercury resistance operon	1850	0	165	12	214
Arsenic resistance	4743	311	584	502	1769
Trehalose Biosynthesis	6794	473	2976	2565	3618
Oxidative stress	11937	78	3279	2091	6741

Others	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Acyclic terpenes utilization	0	0	445	0	22
Alkanesulfonates Utilization	1403	0	2324	0	830
Alkylphosphonate utilization	293	0	952	213	417
Ethylmalonyl-CoA pathway of C2 assimilation	0	0	65	0	21
Exosome	0	186	0	0	3
Formate--tetrahydrofolate ligase (EC 6.3.4.3)	0	0	40	1	124
Gene Transfer Agent	0	0	1048	0	2
Intracellular PHB depolymerase (EC 3.1.1.-)	0	0	57	0	7
Malyl-CoA lyase (EC 4.1.3.24)	0	0	106	0	12
Methylglyoxal Metabolism	2637	555	1611	796	2080
Na+ translocating decarboxylases and related biotin-dependent enzymes	0	0	0	10	63
Nitrate and nitrite ammonification	18464	274	1463	1634	3096
Nitric oxide synthase	627	0	1487	444	1115
Nitrogenase (molybdenum-iron) alpha chain (EC 1.18.6.1)	199	0	6	0	2
Nitrogenase (molybdenum-iron) beta chain (EC 1.18.6.1)	253	0	0	0	0
Nonhomologous End-Joining in Bacteria	0	0	1098	206	168
Photosystem I	6326	0	0	0	757
Photosystem II	9581	0	0	0	1024
Photosystem II-type photosynthetic reaction center	0	0	583	0	21
Plasmid-encoded T-DNA transfer	0	0	2200	10	87
Poly(3-hydroxyalkanoate) depolymerase	0	0	86	0	27
Polyhydroxyalkanoate granule-associated protein PhaF	0	0	11	0	0
Polyhydroxyalkanoic acid synthase	102	0	821	0	325
Ribulose bisphosphate carboxylase large chain (EC 4.1.1.39)	436	0	162	0	19
Ribulose bisphosphate carboxylase small chain (EC 4.1.1.39)	238	0	56	0	8
Soluble cytochromes and functionally related electron carriers	15713	145	1561	3192	1325
Soluble methane monooxygenase (sMMO)	0	0	14	0	155
Succinate dehydrogenase	3558	846	903	587	1330
Tricarboxylate transport system	0	0	1823	0	652

Table 8.3. Abundance of enzymes related to aromatic compounds degradation in the four most abundant taxa found in the metagenome of sample 420. Values: MG-RAST abundances.

Benzoate to catechol	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Benzoate 1,2-dioxygenase (EC 1.14.12.10)	0	0	3	0	3
1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase	0	0	82	0	2
Ortho-halobenzoate 1,2-dioxygenase alpha-ISP protein OhbB	0	0	1	0	4
Ortho-halobenzoate 1,2-dioxygenase beta-ISP protein OhbA	0	0	1	0	0

4-hydroxybenzoate to benzoyl-CoA or protocatechuate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
3-hydroxybenzoate--CoA/4-hydroxybenzoate--CoA ligase	0	0	0	0	5
4-hydroxybenzoyl-CoA reductase, beta subunit (EC 1.3.99.20)	114	0	229	158	605
4-hydroxybenzoyl-CoA reductase, gamma subunit (EC 1.3.99.20)	53	0	258	56	153
putative 4-hydroxybenzoyl-CoA thioesterase	0	0	39	0	29
P-hydroxybenzoate hydroxylase (EC 1.14.13.2)	0	0	109	14	114

Benzoyl-CoA to 3-oxoadipyl-CoA	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Benzoate-CoA ligase (EC 6.2.1.25)	0	0	513	0	75
Benzoyl-CoA oxygenase component A	0	0	9	0	2
Benzoyl-CoA oxygenase component B	0	0	223	0	4
benzoyl-CoA-dihydrodiol lyase (BoxC)	0	0	362	0	5
3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)	0	0	860	0	268

Catechol to beta-ketoadipate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Catechol 1,2-dioxygenase (EC 1.13.11.1)	0	0	227	0	8
Muconate cycloisomerase (EC 5.5.1.1)	1060	0	146	185	67
Muconolactone isomerase (EC 5.3.3.4)	0	0	36	0	19
Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)	39	0	58	0	70

Protocatechuate to beta-ketoadipate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Protocatechuate 3,4-dioxygenase alpha chain (EC 1.13.11.3)	0	0	24	0	26
Protocatechuate 3,4-dioxygenase beta chain (EC 1.13.11.3)	0	0	87	6	90
3-carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2)	0	3	156	0	86
4-carboxymuconolactone decarboxylase (EC 4.1.1.44)	95	0	92	36	288
Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)	39	0	58	0	70
3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase	0	0	7	0	0

Beta-ketoadipate degradation	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)	0	45	189	320	220
Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)	0	45	203	262	273
Beta-ketoadipyl CoA thiolase (EC 2.3.1.-)	0	0	138	0	114

Anaerobic benzoate metabolism	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Benzoate-CoA ligase (EC 6.2.1.25)	0	0	513	0	75
Benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase subunit, BcrC/BadD/HgdB	526	0	0	0	20
Cyclohex-1-ene-1-carboxyl-CoA hydratase (EC 4.2.1.17)	0	0	2	0	0
2-hydroxycyclohexanecarboxyl-CoA dehydrogenase (EC 1.1.1.-)	0	0	2	0	76
2-ketocyclohexanecarboxyl-CoA hydrolase (EC 4.1.3.36)	0	0	0	0	31
Putative pimeloyl-CoA dehydrogenase pimC (large subunit)	0	0	105	0	0
Putative pimeloyl-CoA dehydrogenase pimD (Small subunit)	0	0	58	0	0
Glutaryl-CoA dehydrogenase (EC 1.3.99.7)	0	90	274	802	513
Enoyl-CoA hydratase (EC 4.2.1.17)	36	37	2428	245	619
3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157)	0	58	325	196	172
Acetyl-CoA acetyltransferase (EC 2.3.1.9)	11	126	1362	863	960

Ferulate and lignostilbene to protocatechuate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Feruloyl-CoA synthetase	0	0	371	0	8
lignostilbene-alpha,beta-dioxygenase	166	0	113	0	0
vanillin synthase / trans-feruloyl-CoA hydratase	0	0	29	0	0
Probable VANILLIN dehydrogenase oxidoreductase protein (EC 1.-.-.-)	0	0	15	0	10
Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)	0	0	223	0	22
Vanillate O-demethylase oxygenase subunit (EC 1.14.13.82)	0	0	133	0	22

Chlorogenate to quinate and caffeate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Chlorogenate esterase	0	0	0	0	6

Quinate to protocatechuate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Quinate/shikimate dehydrogenase [Pyrroloquinoline-quinone] (EC 1.1.99.25)	0	0	4	0	11
3-dehydroquinate dehydratase I (EC 4.2.1.10)	0	261	0	0	5
3-dehydroquinate dehydratase II (EC 4.2.1.10)	486	5	148	160	69
3-dehydroshikimate dehydratase	0	64	0	0	0

Caffeate and coumarate to protocatechuate	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Feruloyl-CoA synthetase	0	0	371	0	8
4-coumarate--CoA ligase 1 (EC 6.2.1.12)	93	0	14	0	13
p-hydroxycinnamoyl CoA hydratase/lyase	0	0	0	0	1
vanillin synthase / trans-feruloyl-CoA hydratase	0	0	29	0	0
Probable VANILLIN dehydrogenase oxidoreductase protein (EC 1.-.-.-)	0	0	15	0	10
P-hydroxybenzoate hydroxylase (EC 1.14.13.2)	0	0	109	14	114

Biphenyl degradation	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
Biphenyl-2-3-dioxygenase	0	0	5	0	7
Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)	0	0	229	0	15
cis-2,3-dihydrobiphenyl-2,3-diol dehydrogenase	0	0	0	0	21
2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-)	19	0	60	405	35
2-keto-4-pentenoate hydratase (EC 4.2.1.-)	0	0	94	0	4
4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.-)	0	0	15	0	33

Phenylacetate degradation to beta-ketoadipyl-CoA	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
3-hydroxyacyl-CoA dehydrogenase PaaC (EC 1.1.1.-)	0	0	1	8	26
Phenylacetate degradation enoyl-CoA hydratase PaaA (EC 4.2.1.17)	0	0	11	25	5
Phenylacetate degradation enoyl-CoA hydratase PaaB (EC 4.2.1.17)	0	0	87	30	32
Phenylacetate-CoA oxygenase, PaaG subunit	0	0	199	58	210
Phenylacetate-CoA oxygenase, PaaH subunit	0	0	48	8	48
Phenylacetate-CoA oxygenase, Paal subunit	0	0	126	13	46
Phenylacetate-CoA oxygenase, PaaJ subunit	0	0	87	1	21
Phenylacetate-CoA oxygenase/reductase, PaaK subunit	0	0	172	21	38
Phenylacetate-coenzyme A ligase (EC 6.2.1.30) PaaF	0	0	161	0	27
Phenylacetic acid degradation operon negative regulatory protein PaaX	0	0	138	0	0
Phenylacetic acid degradation protein PaaD, thioesterase	0	0	76	2	7
Phenylacetic acid degradation protein PaaE, ketothiolase	0	0	16	51	25
Phenylacetic acid degradation protein PaaN, ring-opening aldehyde dehydrogenase (EC 1.2.1.3)	0	0	216	88	112
Phenylacetic acid degradation protein PaaY	0	0	1	47	0

4-hydroxyphenylacetate degradation	Cyanobacteria	Ascomycota	Alphaproteobacteria	Bacteroidetes	Others
4-hydroxyphenylacetate 3-monooxygenase, reductase component (EC 1.6.8.-)	0	0	10	0	1
3,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)	0	0	45	0	9
5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60)	0	0	201	39	97
5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)	0	0	51	84	31
5-carboxymethyl-2-oxo-hex-3- ene-1,7-dioate decarboxylase (EC 4.1.1.68)	17	0	131	18	63
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (EC 5.3.3.-)	33	0	131	0	62
2-oxo-hepta-3-ene-1,7-dioic acid hydratase (EC 4.2.-.-)	0	0	155	0	18
2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase (EC 4.1.2.-)	0	9	112	2	39
succinate semialdehyde dehydrogenase (EC 1.2.1.16)	0	40	259	0	412

In addition to this, we, were especially interested in the potential of the microbial community from sample 420 for degrading complex organic matter, particularly lignin derivatives. A step-by-step reconstruction of all the aromatic degradation pathways found in the metagenome of sample 420 can be found in **Table 8.3** and **Figure 8.5**. Central pathways for the degradation of aromatic compounds were reconstructed completely, while key enzymes for several peripheral pathways were supported by the metagenomic data (**Table 8.3**, **Figure 8.5**). Almost all the reads classified as belonging to an enzyme involved in the degradation of aromatic compounds

were assigned to the *Alphaproteobacteria* class. Interestingly, several alternative pathways for the degradation of benzoate were found in the metagenome, including an alternative pathway which does not require an oxygenolytic ring cleavage step, as well as a completely anaerobic pathway (**Table 8.3**, **Figure 8.5**).

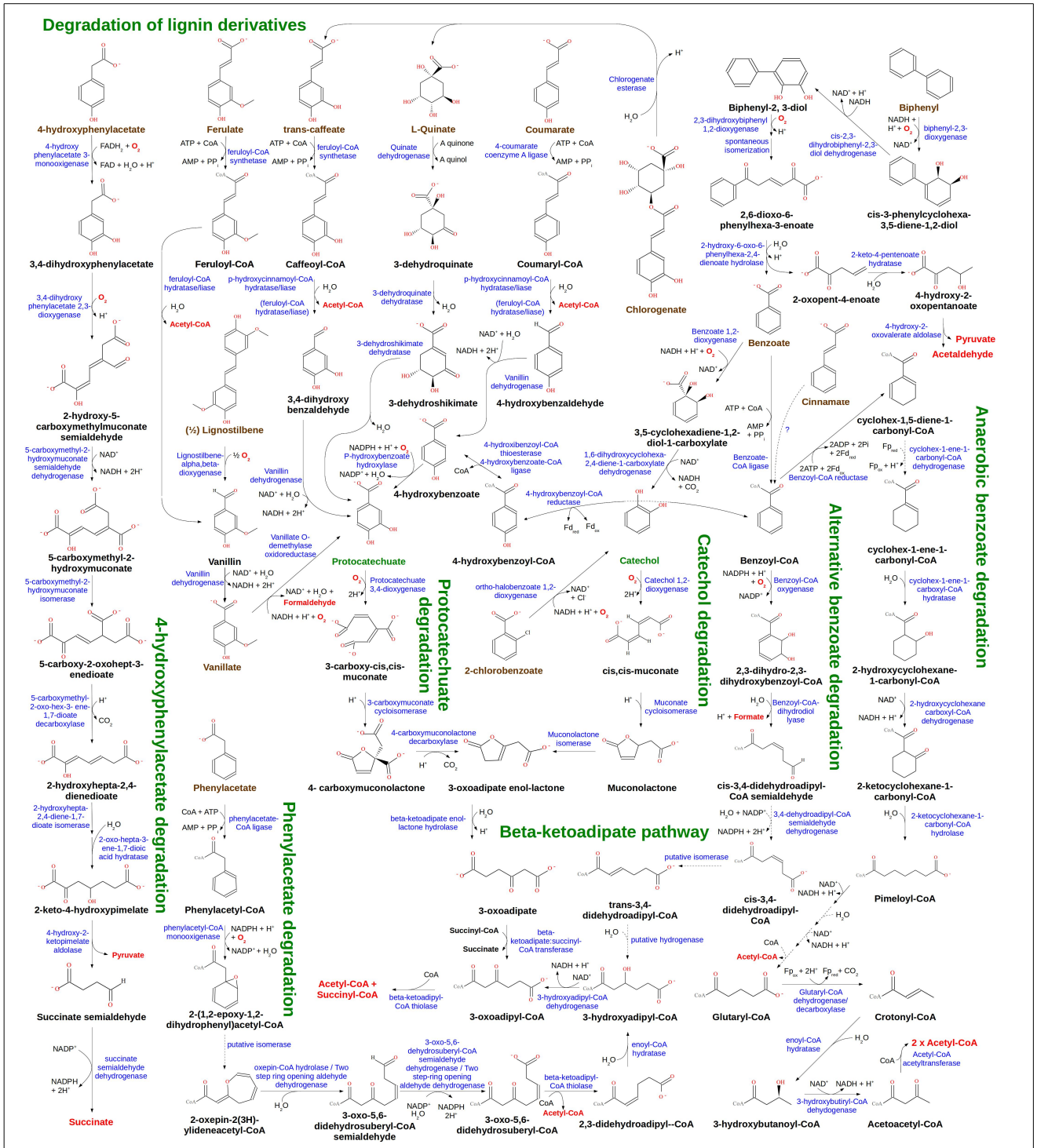


Figure 8.5. Metabolic pathways for the degradation of lignin derivatives and other aromatic compounds found in the metagenome of sample 420. Dashed arrows indicate reactions not directly supported by the metagenomic data.

8.3. Discussion

8.3.1. Assessing potential contamination events during sample retrieval and processing

As described in **section 8.2.2**, the drilling water (DW) control clustered independently from the subsurface samples (**Figure 8.2a**), strongly suggesting that no contamination was introduced during the drilling process. Note that this also covers other contaminants such as shallower sections of the borehole, as their presence inside the core sample would be accompanied by the presence of drilling water.

Our internal laboratory (IC) control, however, did cluster together with two samples (228, 353). Even while taking into account the stringency of our analysis method (**section 8.2.2**) and the fact that both samples were more similar between them than to the IC control, we can not rule out the possibility that a certain degree of contamination was introduced in those two samples during laboratory manipulation.

Finally, the most abundant OTUs were similar to sequences obtained from Epilithic-Dry-Permafrost environments (**Figure 8.2b**), with others being related to sludge, wastewater, contaminated, soil or rhizosphere environments. As expected, OTUs related to water environments were found mainly in the DW control, while some OTUs related to human/animal microbiome were detected cross several samples. The fact that the majority of the OTUs found in the borehole were unrelated to water or microbiome environments could again suggest the absence of contamination. It is worth noting, however, that subsurface organisms are known to be extremely tolerant to radiation, which is highly related to their tolerance to desiccation stress (see **section 1.1.4**). Conversely, it could be argued that DNA from contaminants resistant to radiation-based decontamination processes (such as the one used in this work, see **section 7.2.2**) could have been enriched during our extraction and amplification processes. Furthermore, such organisms could have been incorrectly classified as indigenous endolithic organisms by previous studies using decontamination methodologies similar to our own, thus misleading our annotation-dependent OTU source assignation method. However, if this were to be the case the presence of contaminants would have also shown in the (annotation-independent) hierarchical clustering presented in **Figure 8.2a**. We can therefore conclude safely that sample contamination was not a significant issue during the course of this study.

8.3.2. A versatile endolithic microbial consortium thriving in the deep subsurface of the IPB

The presence of cyanobacterial reads in most of our samples was unexpected, especially after the possibility of sample contamination was discarded (see previous section). Furthermore, the abundance of cyanobacterial subsections II and III was found to be negatively correlated with hydrogen concentrations (**Figure 8.3**), suggesting that their presence might be associated with hydrogenotrophic communities. The existence of endolithic cyanobacteria such as the ones found in this study (**section 8.2.4**) has been known for long (Friedmann, 1980; Knoll *et al.*, 1986; Friedmann & Meyer 1986), and their presence in deep subsurface environments has been reported several times (Kormas *et al.*, 2003; Newberry *et al.*, 2004; Rastogi *et al.*, 2010); however, to the best of our knowledge their potential role in such ecosystems has never been discussed. Several dark metabolisms, such as different types of fermentation, coupled in some cases to non-respiratory sulphur reduction, have been described in *Cyanobacteria* (see Stal, 2012 for a thorough review). In general they are supposed to only able to degrade reserve polysaccharides previously obtained by photosynthesis (Oren & Shilo, 1979; Stal & Moezelaar, 1997; Troshina *et al.*, 2002), but at least one case exists in which anaerobic uptake and utilization of extracellular carbohydrates has been described (Bagchi *et al.*, 1990). Furthermore, during the writing of this thesis, Stuart *et al.* (2015) also reported the ability of some cyanobacteria to reuse extracellular organic carbon in microbial mats.

In order to elucidate the metabolic strategies of the microbial community present in BH10, as well as the potential roles of cyanobacteria in deep subsurface environments, we chose to perform shotgun metagenomic sequencing on sample 420, as our results (**Figure 8.2c**; **Figure 8.3**) indicated that it contained a relatively diverse microbial community in which the presence of cyanobacteria was correlated to a decrease in hydrogen concentration. Furthermore, organic carbon and intact fatty acids were found in sample 420, indicating the presence of an active microbiota with access to buried organic matter, while the presence of fractures at that area suggested that water and nutrient input might be higher than in other depths of the borehole.

According to our metagenomic data, the bacterial community found in sample 420 was dominated by cyanobacteria, alphaproteobacteria and bacteroidetes. Ascomycotal reads were also highly abundant (**Figure 8.4**). While the taxonomic classification of the metagenomic reads might suggest a broader diversity, the 16S data leads us to believe that each of the bacterial groups is mainly represented by a single genus (**Annex 3 - Table S3**). Other taxonomic groups such as *Actinobacteria*, *Deinococcus-Thermus* and *Gammaproteobacteria* were also present in the

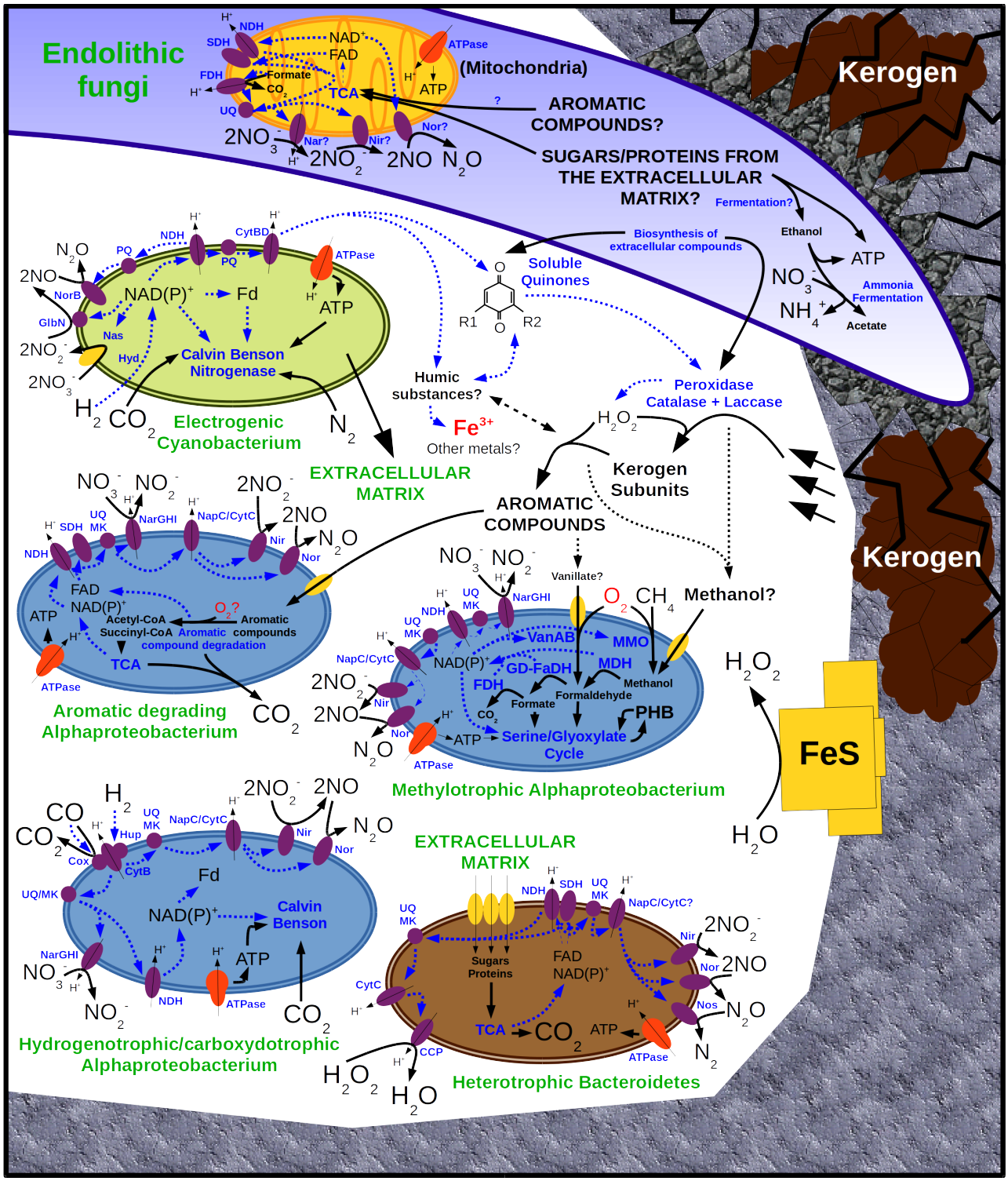


Figure 8.6. Metabolic model of the microbial community found in sample 420. Names in blue correspond to enzymes, electron carriers or metabolic pathways; names in black correspond to organic and inorganic compounds. Dotted blue arrows correspond to electron flow; black arrows correspond to mass flow. Dotted black arrows or names with interrogations indicate processes that have been described in the literature but are not directly supported by the metagenomic data. **Abbreviations: ATPase**, ATP synthase; **CCP**, cytochrome C peroxidase; **Cox**, carbon monoxide dehydrogenase; **CytBD**, cytochrome bd; **CytC**, cytochrome C; **FDH**, formate dehydrogenase; **GlbN**, cyanoglobin; **GD-FaDH**, glutathione-dependent formaldehyde dehydrogenase; **Hup**, uptake hydrogenase; **Hyd**, NAD+ reducing hydrogenase; **MDH**, methanol dehydrogenase; **MK**, menaquinone; **MMO**, methane monooxigenase; **NapC/CytC**, denitrification associated cytochrome C; **Nar**, respiratory nitrate reductase; **Nas**, assimilatory nitrate reductase; NDH, NADH dehydrogenase; **Nir**, respiratoty nitrite reductase; **Nor**, respiratory nitric oxide reductase; **Nos**, nitrous oxide reductase; **PQ**, plastoquinone; **SDH**, succinate dehydrogenase; **TCA**, tricarboxylic acid cycle; **UQ**, ubiquinone; **VanAB**, vanillate o-demethylase oxidoreductase.

metagenome, but their small abundance precluded a more detailed analysis of their potential metabolic roles. The community was found capable of resisting a wide array of environmental stresses such as desiccation, oxidative stress, salt stress or metal toxicity (**Table 8.2**), while having the potential to perform several different metabolisms (**Table 8.3**; **Figure 8.5)**: hydrogenotrophy (*Alphaproteobacteria*, *Cyanobacteria*), carboxidotrophy (*Alphaproteobacteria*), methylotrophy (*Alphaproteobacteria*), heterotrophy (Ascomycota, *Bacteroidetes*), and degradation of refractory organic matter (*Ascomycota*, *Alphaproteobacteria*). Based on these results, we propose that these organisms are associated together to form a metabolically versatile endolithic holobiont (**Figure 8.6**), in which each partner provides essential functions for the survival of the community.

In order to thrive in deep subsurface environments, organisms must gain access to water and nutrients. Given the non-porous nature of the silicates found in BH10, this likely requires the ability of burrowing into rocks in order to unveil buried resources and increase the available space for colonization. Note that, even when below the water table, endolithic organisms can be water-limited if they have no available pore space; conversely, biological rock weathering will lead to additional circulation of water and nutrients. Deep subsurface microorganisms also need to be able to exploit the different biotic and abiotic energy sources available to them (see **section 1.1.2**). The heterogeneous nature of the continental subsurface (see section **1.2.3**) makes metabolic versatility a valuable treat in order to colonize different neighbouring microenvironments. Finally, microorganisms need to be able to survive with the very low energy fluxes characteristic of deep subsurface ecosystems, while withstanding a wide array of environmental stresses. In the following paragraphs we will discuss how the different partners in the holobiont could contribute towards the achievement of these goals.

8.3.2.1. Endolithic Ascomycota and the potential for kerogen degradation

Fungi are well-known for their ability to form complex symbiotic associations, in which they provide water, nutrients and colonizable niches to their bacterial partners in exchange for fixed nitrogen and carbon (see Grube & Berg, 2009 for a recent review). The fungal metagenomic reads obtained from sample 420 were further classified as belonging to the *Eurotiomycetes*, *Sordariomycetes* and *Dothideomycetes* classes, which are known to harbour many extremotolerant rock-inhabiting species (Ruibal *et al.*, 2009; Nai, 2014; Muggia *et al.*, 2015). Those non-lichenized rock inhabiting ascomycetous black fungi are thought to have an ancient origin, being the ancestors of both pathogenic fungi and mutualistic lichens (Gueidan *et al.*, 2008; Gueidan *et al.*, 2011). They

are able to penetrate igneous hard rocks, like granite and basalt as well as sedimentary soft rock such as limestone, sandstone or marble (Sterflinger & Krumbein, 1997; Sterflinger, 2000). Nutritionally, they can rely exclusively on sparse, airborne, low molecular weight nutrients (Gueidan *et al.*, 2008; Gostinčar *et al.*, 2012), although some of them can also form mutualistic interactions with photobionts (Gorbushina *et al.*, 2005; Muggia *et al.*, 2013) or degrade aromatic hydrocarbons (Gostinčar *et al.*, 2012).

Unfortunately, the metabolic reconstruction of the fungal population present in sample 420 was incomplete, probably due to the higher size of eukaryotic genomes and/or to the presence of a consortia of different lichenized and non-lichenized fungal species (Muggia *et al.*, 2013). The available carbon and energy sources for the fungi present in sample 420 would be either small organic compounds (sugars, organic acids; see **Figure 8.1**) released by their autotrophic partners (see below), or buried kerogen. Some black fungi have been hypothesised to be capable of lignin and refractory organic matter degradation (Nai *et al.*, 2013; Kietäväinen & Purkamo, 2015). The refractory organic matter present in the IPB subsurface is mainly type III kerogen (Barberes, 2015) which is derived from lignic debris (Vandenbroucke & Largeau, 2007). Lignin depolymerization and the subsequent degradation of lignin monomers could be therefore good model processes for assessing the refractory organic matter degradation potential in the deep subsurface of the IPB.

Many fungal species (especially basidiomycetes, but also ascomycetes) are able to degrade lignin via extracellular oxidative enzymes such as laccases or peroxidases (see Bugg *et al.*, 2011 for a thorough review on lignin degradation pathways). These enzymes use either oxygen or hydrogen peroxide to oxidatively cleave lignin or other complex organic molecules. Fungal peroxidases and multicopper-oxidases (of which catalase is a representative) were indeed detected in the metagenome of sample 420 (**Table 8.2**). Another possible path for lignin degradation involves the generation of reactive oxygen species via extracellular quinone-driven Fenton chemistry (Guillén *et al.*, 1997; Keren *et al.*, 1999; Bugg *et al.*, 2011). Interestingly, fungal quinate permeases were found in the metagenome (**Table 8.2**). While they are usually involved in quinate degradation (with quinate being an intermediate in lignin synthesis and breakdown), they could also allow the excretion of other structurally related polar compounds.

These two paths for lignin oxidation ultimately require molecular oxygen, and oxygenated water from shallower regions of the borehole could indeed be entering the system through the fractures detected in the area surrounding sample 420. However, anaerobic lignin degradation linked to the presence multicopper oxidases and peroxidases has also been described (van Waasbergen *et al.*, 2000). We hypothesize that oxidative kerogen breakdown could be occurring in the deep subsurface

of the IPB even in the absence of an oxygenated water influx. Water can react with defect surface sites on pyrite to ultimately produce hydrogen peroxide, even under dark, anoxic conditions (Borda *et al.*, 2001; 2003). Its production by pyrite is dependent on the exposed crystal area, and is not limited to the initial contact with water (Borda *et al.*, 2001). This hydrogen peroxide could be used by extracellular peroxidases – or a combination of catalases and oxidases – to break down refractory organic compounds. The ability of endolithic fungi to actively penetrate in hard rocks would increase the access of the community to buried carbon sources, while also exposing more pyrite surfaces to water, with the subsequent release of hydrogen peroxide (**Figure 8.6**).

According to our results, it is unlikely that the fungi is able to fully degrade refractory organic matter by itself, as it appears to lack most of the pathways for aromatic compound degradation. Instead, we propose that it contributes to the initial unveiling and oxidative breakdown of kerogen as described above, while the resulting aromatic molecules are degraded by an alphaproteobacterium (see below). In this scenario, the ascomycetous fungi would feed on small organic compounds provided by its bacterial partners (**Figure 8.6)**. Assuming that not enough oxygen is available to sustain fully aerobic metabolisms, the fungi could obtain energy via fermentation (**Table 8.2**). Ammonia fermentation couples oxidation of ethanol to non-respiratory nitrate reduction (Takasaki *et al.*, 2004). Several fungi has been shown to perform respiratory denitrification (Shoun *et al.*, 1990; Kobayashi *et al.*, 1996), generally associated to formate oxidation (Kuwazaki et al 2003; Ma *et al.*, 2008). While a fungal formate dehydrogenase was found in the metagenome of sample 420, there were no evidences of fungal enzymes for denitrification. As discussed above, we can not rule out their presence completely due to the low coverage of the fungal part of the metagenome.

8.3.2.2. Endolithic cyanobacteria and biofilm formation

As discussed above, the presence of cyanobacteria has been reported in other deep subsurface environments, but to the best of our knowledge their potential roles in such dark and ecosystems have not been addressed so far. Our physicochemical, taxonomic and metagenomic results (**Figure 8.3**; **Table 8.2**) showed that cyanobacteria are abundant in the subsurface of the BH10 area, and suggest that, at least in sample 420, they might be able to obtain energy by oxidizing molecular hydrogen (**Figure 8.6**).

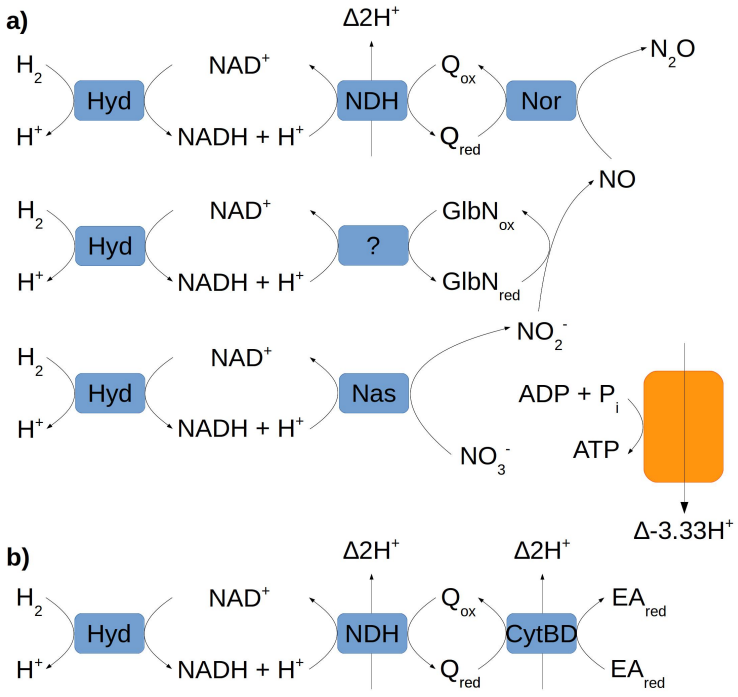
Hydrogenases are widespread in cyanobacteria, in which they are used either to minimize energy losses during nitrogen fixation (oxidizing molecular hydrogen and allocating the resulting electrons

to the plastoquinone pool) or to dispose of excess reducing power (by reducing protons and releasing hydrogen) (Bothe *et al.*, 2011; Stal, 2012). We found a cyanobacterial NADH evolving hydrogenase (**Table 8.2**; **Annex 3 - Table S4**) in the metagenome of sample 420. NADH could be converted to NADPH or ferredoxin via NAD(P) transhydrogenase and Ferredoxin-NADP(+) reductase in order to provide reducing power for biosynthetic pathways. NADH could also be oxidised by NADH dehydrogenase (NDH), generating proton-motive force (PMF) and transferring electrons to plastoquinone (**Figure 8.6)**. NDH is part of the standard cyanobacterial electron transport chain, in which the reduced plastoquinone is usually reoxidised by cytochrome b6f (see Vermaas, 2001 for a simple review). However, cytochrome b6f derives electrons via plastocyanin to either photosystem I or an oxygen-consuming terminal oxidase. While the genes for both pathways were detected in the cyanobacterial subset of the metagenome, it is unlikely that they are operative in the dark, potentially anoxic deep subsurface of the IPB. Based on our metagenomic results and on the existing literature, we are able to propose two putative electron chains that would couple hydrogen oxidation to energy transduction in the conditions found in the BH10 subsurface (**Figure 8.6**).

A nitric oxide reductase (*nor*) was found in the cyanobacterial fraction of our metagenome (**Table 8.2**). Cyanobacterial nitric oxide reductases connected to an electron transport chain have been described by previous studies (Büsch *et al.*, 2002). This enzyme reduces nitric oxide (NO) by accepting electrons from reduced quinones, albeit it is unable to oxidise cytochrome c or generate PMF. Some *Cyanobacteria* are able to generate NO under anoxic conditions via nitrite reduction with cyanoglobin (Mallick *et al.*, 1999; Sturms *et al.*, 2011), which was also found in our metagenome. Nitrite could be obtained directly from the environment, or by an assimilatory nitrate reductase (Mallick *et al.*, 1999). In order for this reaction to progress, the resulting oxidised cyanoglobin must be reduced again, and NO must be quickly scavenged so as not to inactivate the reduced cyanoglobin (Sturms *et al.*, 2011). Since cyanoglobin is located in the membrane, it's co-localization with NOR would minimize cyanoglobin inhibition by NO. The *in-vivo* mechanism for reducing oxidised cyanoglobin is yet unknown (Sturms *et al.*, 2011). If cyanoglobin was to be reduced by NADH (as is the case for methemoglobin), only one in three hydrogen molecules consumed would result in PMF generation, and PMF would only be generated by NDH (**Figure 8.7a**). Assuming that NDH translocates 2 protons per NADH oxidised, and that 3.33 protons are required for synthesizing 1 ATP molecule, this mechanism would yield 0.2 of mol ATP per mol of hydrogen consumed, consuming 0.33 mol of nitrate in the process. This would be still more efficient than, for example, hydrogenotrophic methanogenesis, which yields 0.12 mol ATP per mol

of hydrogen consumed (as calculated by Kaster *et al.*, 2011), and the yield would drastically improve if nitrite or NO were available instead of nitrate. In any case, let the reader note that the calculations presented here are rough estimates. While the required enzymes were indeed detected in the metagenome (**Table 8.2**), a detailed thermodynamic investigation would be needed in order to determine the actual feasibility of the proposed pathway.

Figure 8.7. Putative cyanobacterial dark energy transduction mechanisms.



Putative electron transfer chains based on the results of the cyanobacterial part of the metagenome of sample 420, as well as on the existing literature.
a) Nitrate reducing.
b) Electrogenic (see **Figure 8.6** for a sketch depicting the potential extracellular electron sinks).
Abbreviations: Hyd, NADH evolving hydrogenase; **NDH**, NADH dehydrogenase; **Q**, quinone; **Nor**, nitric oxide reductase; **GlbN**, cyanoglobin; **Nar**, assimilatory nitrate reductase; **CytBD**, cytochrome BD; **EA**, extracellular electron acceptor. ATPase (orange box) is depicted to show the standard ATP gain per translocated proton.

Another more efficient energy transduction mechanism would involve the transfer of electrons to extracellular acceptors. This has been observed in several cyanobacterial genera (Zou *et al.*, 2009; Pisciotta *et al.*, 2010; Hasan *et al.*, 2014; Nishio *et al.*, 2015) including the *Leptolyngbya* genus (Pisciotta *et al.*, 2010; Hasan *et al.*, 2014), which was found in sample 420 (**Table 8.1**; **Annex 3 - Table S3**). This process is thought to regulate the redox state of plastoquinone by shedding excess energy to the environment when it becomes over reduced in high light (Pisciotta *et al.*, 2010). Electron transfer from hydrogen to plastoquinone via hydrogenase and NDH (**Figure 8.7b**) would also result in an excess of reduced plastoquinol, since cytochrome b6f, its usual electron acceptor (Vermaas, 2001; Trubitsin *et al.*, 2005), would also be in its reduced state due to the absence of both light and oxygen. Pisciotta *et al.* (2011) showed that, in the absence of cytochrome b6f activity, plastoquinol is instead reduced by a cytochrome bd quinol oxidase, which is in turn responsible for electron transfer to the environment. We found reads similar to cytochrome D/bd ubiquinol oxidase (*cydAB*) in the cyanobacterial fraction of the metagenome (**Table 8.2**). It is not yet clear if bd quinol

oxidase transfers electrons directly to extracellular electron acceptors or if additional downstream components are involved (Pisciotta et al 2011), however the fact that natural *Leptolyngbya* biofilm are capable of electrogenic activity (Pisciotta *et al.*, 2010) suggests that this activity is performed by some cyanobacteria under *in vivo* conditions.

Under the same assumptions as described for the putative nitrous oxide reducing electron transport chain, hydrogen oxidation coupled to extracellular acceptor reduction would yield 1.2 mol ATP per mol of hydrogen consumed. The ultimate fate of these extracellular electrons can not be elucidated solely from our metagenomic data, however several pathways would be possible according to our model for the sample 420 ecosystem (**Figure 8.6**). They would involve an initial electron transfer to extracellular quinones, which were also proposed as participants in kerogen degradation (see previous section). Extracellular quinones have been found to be mediators to other final acceptors (Newman & Kolter, 2000; Yamazaki *et al.*, 2002); these quinones can be produced by microorganisms, but can also be present in the environment, such as the quinone moieties present in humic acids (Kappler *et al.*, 2004). Humic acids are indeed expected to appear during the breakdown of the type III kerogen present in the deep subsurface of the IPB (Sara Lincoln, personal communication). These intermediate compounds would donate their electrons to several possible ultimate acceptors. Reduced quinones are substrates for fungal extracellular peroxidases (Valli *et al.*, 1991; Valli *et al.*, 1992). In this case, hydrogen peroxide would be the final electron acceptor. Similarly, the Fenton chemistry-based lignin degradation pathway discussed in previous section would involve transfer of electrons from reduced quinones to hydrogen peroxide, with iron acting as an intermediate, and the resulting hydroxyl radical attacking the lignin/kerogen molecules. This process would result in additional oxidised quinone moieties, which can be in turn used as additional electron acceptors (DeAngelis *et al.*, 2013). This is especially interesting from the holobiont perspective, as it links the cyanobacterial energy transduction pathway to fungal kerogen degradation. Finally, humic acids also can shuttle electrons to poorly soluble oxidised iron or manganese minerals (Kappler *et al.*, 2004).

By coupling hydrogenotrophy with the reduction of either nitrogen compounds or extracellular electron acceptors, the cyanobacterium found in sample 420 would be able to act as a primary producer for the ecosystem, fixing carbon via the Calvin-Benson cycle and releasing part of as an extracellular polysaccharide from which the rest of the members of the community can benefit. This cyanobacterium would also potentially be able to fix nitrogen, since a cyanobacterial nitrogenase was found in the metagenome of sample 420 (**Table 8.2**). In practice, it is unlikely that this would be happening, as assimilatory reduction of the nitrate present in the sample (**Figure 8.1**) would

require less energy. The cyanobacterial sheath is probably playing a key role in the maintenance of the holobiont: cyanobacterial EPS contributes significantly to the structure of known endolithic biofilms and their interaction with the mineral substrate (Barker & Banfield, 1996). EPS matrices contribute to rock weathering by imbibing water, increasing their volume and causing rocks to crack (Ascaso *et al.*, 2002), but also by trapping particles in the vicinity of the cells (Wolfaardt *et al.*, 1999; Gorbushina *et al.*, 2000). In addition to direct physical damage, they are also involved in chemical biomobilization and biomineralization processes (Barker & Banfield, 1996; Barker *et al.*, 1998). Through these mechanisms, endolithic cyanobacteria increase the available pore space, and facilitate its colonization by other organisms, including heterotrophs (Blackhurst *et al.*, 2005; de los Ríos *et al.*, 2007). Under our model (**Figure 8.6)**, the cyanobacterium would be the main contributor to biofilm formation (see Rossi & De Philippis, 2015 for a thorough review on the roles of cyanobacteria in mixed biofilms), providing heterotrophic organisms with low molecular weight compounds that can be used as either carbon sources or protectors against desiccation.

8.3.2.3. Aromatic compound-degrading alphaproteobacteria

The *Alphaproteobacteria* class is ecologically versatile, harbouring species adapted to a broad range of environments and lifestyles. Many of them are able to stablish trophic interactions with other organisms, ranging from the highly specific endosymbiosis found in legume nodules to a wide array of less specific and ecologically heterogeneous associations (Pini *et al.*, 2011). This includes their presence in fungal-bacterial communities such as lichens or ectomycorrhizal symbionts (Barbieri *et al.*, 2007; Grube & Berg, 2009).

According to the 16S results, the majority of the alphaproteobacterial population belonged to a single *Bradirhizobium* species closely related to *B. elkanii* (**Table 8.1**; **Annex 3 - Table S3**), which has been also found to be the most representative species associated with some ascomycetous fungi such as *T. magnatum* (Barbieri *et al.*, 2007). This was consistent with the metagenomic results, which showed a large proportion of reads classified as belonging the *Rhizobiales* order in the alphaproteobacterial fraction of the community. The alphaproteobacteria found in sample 420 were metabolically diverse, being potentially capable of growing via hydrogenotrophy, carboxydotrophy, methylotrophy, or by degrading aromatic compounds. Nitrate was found to be the most likely electron acceptor under anaerobic conditions (**Table 8.2**; **Figure 8.6**).

Hydrogenotrophy and carboxydotrophy frequently appear together in alphaproteobacteria (Meyer & Schlegel, 1983; Lorite *et al.*, 2000). They have been shown to occur aerobically, but also coupled to

nitrate reduction (Neal *et al.*, 1983; King, 2006), albeit nitrate-dependent anaerobic CO oxidation is thought to only provide energy for maintenance metabolism (King, 2006). Since both the uptake hydrogenase and the aerobic carbon monoxide dehydrogenase are membrane bound and transfer electrons to the quinone pool (Bernhard *et al.*, 1997; Wilcoxen *et al.*, 2011), under a purely hydrogenotrophic/carboxidotrophic metabolism NADH would have to be produced via reverse electron flow. The Calvin-Benson cycle was also detected in the alphaproteobacterial fraction of the metagenome, meaning that this organism would be potentially able to grow autotrophically.

Almost-complete pathways for the degradation of lignin derivatives and other aromatic compounds were found in the alphaproteobacterial fraction of our metagenome (**Table 8.3**; **Figure 8.5**). Benzoate and protocatechuate are central intermediates in the degradation of aromatic compounds, and their degradation usually requires oxygen for ring hydroxylation and cleavage, regardless of the actual electron acceptor used in respiration. Three different pathways for benzoate degradation were present: the canonical catechol pathway, requiring two molecules of oxygen per molecule of benzoate degraded; the aerobic *Azoarcus* pathway which includes a non-oxygenolytic ring cleavage step and thus consumes only one oxygen molecule per molecule of benzoate (Gescher *et al.*, 2005); and the anaerobic *Azoarcus* pathway (Barragán *et al.*, 2004). The two *Azoarcus* pathways were incomplete in our metagenome, albeit key enzymes exclusive of both of them, such as benzoyl-CoA-dihydrodiol lyase or cyclohex-1-ene-1-carboxyl-CoA hydratase, were detected (**Table 8.3**; **Figure 8.5**). Regarding protocatechuate degradation, only the canonical pathway (consuming one molecule of oxygen per molecule of protocatechuate degraded) was found. It has been suggested that anaerobic protocatechuate degradation to glutaryl-CoA might proceed via 3-hydroxybenzoyl-CoA using enzymes from the anaerobic benzoate degradation pathway (Laempe *et al.*, 2001; Philipp *et al.*, 2002). To the best of our knowledge, the complete anaerobic protocatechuate degradation pathway has not yet been fully elucidated. Nonetheless, it has been shown that some of the participant enzymes are actually promiscuous enzymes from other pathways (Ding *et al.*, 2008). For example, protocatechuate conversion to 3-hydroxy-1,5-cyclohexadien-1-carbonyl-CoA is carried out by the non-specific enzymes 3-hydroxybenzoate-CoA ligase, 4-hydroxybenzoyl-CoA reductase and benzoyl-CoA reductase (Ding *et al.*, 2008), all of which were found in our metagenome (**Table 8.3**).

These results suggest that the alphaproteobacterium found in the metagenome of sample 420 possesses alternative pathways for aromatic hydrocarbon degradation that could be active under different oxygen concentrations. In addition to the potential for completely anaerobic pathways, the presence of trace amounts of oxygen (originating either by the infiltration of oxygenated waters

from above, or by the action of peroxidases over pyrite-derived hydrogen peroxide) could support the more energetically favourable oxygenolitic breakdown of aromatic rings, with the resulting compounds being catabolysed anaerobically (Lucey & Leadbetter, 2014). Nitrate is a known electron acceptor in anaerobic aromatic compound degradation (Barragán *et al.*, 2004).

The ability to grow on methanol is also relevant, as it can be a by-product of lignin degradation (Ander *et al.*, 1985; Ander & Eriksson, 1985). The alphaproteobacterium present in sample 420 had all the enzymes required for methanol oxidation to CO_2, as well as malyl-CoA synthase, a key enzyme in the serine-glyoxylate cycle for the assimilation of C1 compounds (**Annex 4 – Figure S2**). Genes for polyhydroxybutyrate (PHB) metabolism were also found (**Table 8.2**). Most methylotrophs lack isocitrate lyase, and therefore use a modified version of the serine-glyoxylate cycle in which glyoxylate is regenerated via a pathway involving enzymes and intermediates from the polyhydroxybutyrate cycle (Korotkova *et al.*, 2002; **Annex 4 – Figure S2**). The enzymes for both pathways were present in our metagenome (**Annex 3 - Table S4**). The complete glyoxylate cycle is also used by isocitrate lyase-positive organisms to generate metabolic intermediates from acetyl-CoA, which is one of the main products of aromatic hydrocarbon degradation (**Figure 8.5**). It is therefore possible that the co-existence of both the glyoxylate shunt and the PHB pathway for glyoxylate regeneration indicates the presence of two different alphaproteobacterial species: a hydrocarbon degrader using the glyoxylate cycle for biosynthesis and a methylotroph using the serine-PHB cycle for C1 assimilation.

Finally, a soluble methane monooxigenase was found in the alphaproteobacterial metagenome (**Table 8.2**). If small amounts of oxygen were present in sample 420 (see above), the methylotrophic alphaproteobacterium would also be able of oxidizing methane to methanol. Methanol oxidation could then be coupled to nitrate reduction in order to obtain energy and fixed carbon (**Figure 8.6**). The levels of methane in sample 420 are low when compared to other depths of the borehole (**Figure 8.1**); redundancy analysis, however, appears to indicate that neither sample 420 nor the *Alphaproteobacteria* class are significant methane sinks in the BH10 system (**Figure 8.3**).

Together, our results indicate that the alphaproteobacterial community found in sample 420 is linked with the breakdown of refractory organic matter, with a predominant species degrading kerogen monomers, and a less abundant species consuming the methanol generated during kerogen breakdown. The fact that alphaproteobacterial peroxidases and oxidases were also detected (**Table 8.2**) raises the question of whether the fungi is actually required for kerogen degradation. Alphaproteobacteria have been proposed to be one of the main contributors to lignin degradation in anoxic soils instead of fungi (DeAngelis *et al.*, 2011), while poplar wood has been shown to be

anaerobically degraded by a microbial consortium composed of members from *Alphaproteobacteria*, *Bacteroidetes* and *Cyanobacteria*, among others (van der Lelie *et al.*, 2012). For example, *Phanaerochaete sordida*, one of the most efficient known lignin-degrading fungi, is deeply associated with a bacterial partner to the point of being unable to grow in axenic cultures (Lim et al., 2003). This suggests that, at least in some cases, the degradation of refractory organic matter proceeds through a physiologically obligate – at least for the fungi – syntrophic interaction, which strengthens our deep subsurface holobiont hypothesis. In any case, regardless of whether they actually participate in kerogen breakdown or not, the ability of the fungal lineages found in sample 420 to actively bore into rocks would result in their hydrocarbon-degrading alphaproteobacterial partners having increased access to buried kerogen deposits.

8.3.2.3. Biopolymer-degrading bacteroidetes

Bacteroidetes are thought to be specialized in degrading complex organic matter, especially in the form of biopolymers such as polysaccharides and proteins (Thomas *et al.*, 2011). They have been found in a wide array of environments, including anaerobic lignin degradation consortia (van der Lelie *et al.*, 2012), in which they are key biomass degraders.

Our results suggest that the bacteroidetes found in sample 420 is playing a similar role, as it possessed a high abundance of genes related to sugar and peptide uptake and catabolism (**Table 8.2**). This figure is especially relevant given its relatively low abundance in the community. According to our metagenome, these bacteroidetes would be adept at degrading polysaccharides containing mannose, rhamnose, fucose, xanthosine, galacturonic acid and N-acetylglucosamine, using nitrate as the electron acceptor (**Table 8.2**) under anaerobic conditions. Mass spectrometry analysis confirmed the presence of some of these sugars in the BH10 subsurface (Victor Parro, personal communication). A sodium-translocating NADH-quinone oxidoreducetase was found in the bacteroidetes fraction of the metagenome (**Table 8.2**). Using sodium instead of protons for building a transmembrane gradient minimizes energy loss due to ion leakage (Vossenberg *et al.*, 1995), which might be relevant under slow metabolic conditions. Finally, the presence of a cytochrome-c peroxidase (**Table 8.2**) would also allow for hydrogen peroxide to be used as the terminal acceptor.

While in the lignin-degrading consortium described by van der Lelie *et al.*, (2012) the bacteroidetes were thought to degrade the polysaccharidic fraction of lignocellulose, it is unlikely that this is happening for the kerogen in the deep subsurface of the IPB, as polysaccharides would have been

destroyed during early diagenesis (Mukhopadhyay, 1992). Instead, we propose that they are are feeding on cyanobacterial and fungal extracellular components (see Thomas *et al.*, 2011). Cyanobacterial extracellular polysaccharide (EPS) consumption by bacteroidetes has been reported in microbial mats (Cole *et al.*, 2014). EPS in natural biofilms represent a more refractory, partially degraded, remnant of the original polymer molecules, with a variable equilibrium between EPS production and consumption in the different biofilm layers (Decho *et al.*, 2005). While it is possible that the role of bacteroidetes is merely commensalistic, limited EPS degradation could also have a positive effect in the community by controlling biofilm permeability, maturation stage, and making organic matter available to other, less versatile heterotrophs in the community.

8.3.3. Microbial processes operating in the BH10 subsurface

Unlike in BH8 (see **section 5.3.2**), no clear trends in microbial composition appeared with depth, apart from a somewhat higher abundance of actinobacteria at the shallower samples, and a higher abundance of cyanobacteria at the deepest part of the borehole (**Figure 8.2c**). The fact that the rocks in borehole BH10 were much less fractured than the ones in borehole BH8 (**section 3.4.3**) suggests that the presence of local fissures, which imply access to water and nutrients, might be a more relevant factor than depth or bulk mineralogy in determining the composition and activity of the BH10 subsurface microbiota. Consistently with this hypothesis, higher fatty acid levels were detected in the fault zones around 350 and 420 mbsf, as well as in the more fractured Culm zone beyond 500 mbsf (**Figure 8.1**). Total organic carbon showed a single peak at 420 mbsf; however, it probably represents buried organic compounds rather than extant biological materials (see **section 1.1.2** and previous section).

In spite of the presence of pyrite, microorganisms related to the iron and sulphur cycles were not especially abundant in BH10 (**Table 8.1**), with the exception of a potentially iron-oxidizing *Acidovorax*. Instead, the most abundant community members either belonged to the *Cyanobacteria* phylum, or to ubiquitous soil groups such as *Arthrobacter* (*Actinobacteria*), *Bradyrhizobium* (*Alphaproteobacteria*) or *Sediminibacterium* (*Bacteroidetes*). As discussed in the above sections, these results are not due to contamination during sample retrieval or processing, and instead truly represent the microbial community composition at several depths of the BH10 borehole. The absence of pyrite oxidizing microorganisms such as *Leptospirillum* or *Acidithiobacillus* can be attributed to a low oxygen availability in the saturated zone of the aquifer. It has been described that pyrite oxidation does not proceed under a water table unless there is a high input of oxygen-saturated water (Hammack & Watzlaf, 1990), which is unlikely to happen in the unfractured rocks

of BH10. The absence of pyrite oxidation would also explain the much lower sulphate concentrations (ranging from 100 to 0 ppm) present in BH10 when compared to BH8 (two orders of magnitude larger, see **Figure 8.1**; **section 5.2.1**). Nonetheless, pyrite might be playing a different role in the BH10 ecosystem, by generating hydroxyl radicals and hydrogen peroxide even under dark, anoxic conditions (Borda *et al.*, 2001; 2003). This might be able to support oxidative metabolisms to a certain extent (Davila *et al.*, 2008), albeit the experimental validation of this hypothesis is beyond the scope of this thesis. Nitrate, another potential electron acceptor for anaerobic respiration, was present in somewhat similar concentrations in both boreholes (ca. 1.5 ppm). In general, the Tinto-Odiel aquifer has relatively high nitrate concentrations, which are attributed to the use of fertilizers for intensive farming in the area (Porras-Martín *et al.*, 1985; Cavaller *et al.*, 2005). Sulphate reduction can be inhibited by nitrate, and nitrate reducers outcompete sulphate reducers under electron donor limitations (Achtnich *et al.*, 1995; Percheron *et al.*, 1999). This is likely occurring in the BH10 area, as nitrate reducing bacteria vastly outnumbered sulphate reducers (**Annex 3 - Table S3**), and suggest that nitrate, entering the Peña de Hierro area via groundwater flow, is the primary electron acceptor in the anaerobic parts of the BH10 system.

Hydrogen, methane and soluble ferric ion concentrations were found to significantly influence microbial composition throughout the BH10 borehole (**Figure 8.3**). The fact that ferric iron and methane concentrations were correlated can be interpreted as either iron reduction inhibiting methanogenesis (Lovley & Phillips, 1987; Van Bodegom *et al.*, 2004; Zhang et al 2009) or the presence of ferric iron-dependent anaerobic methane oxidation (Amos *et al.*, 2012). Both processes would require the presence of an archaea (performing either standard or reverse methanogenesis) and an iron-reducing bacteria. As noted in the results section, even though methanogenic and methanotrophic archaea were indeed found in BH8 (see **section 5.2.3**), the PCR primers used in BH10 did not allow us to target the archaeal diversity. At any rate, their presence in BH10 would not be surprising, as both methanogenesis and anaerobic methane oxidation have been previously described in deep subsurface environments (Chapelle *et al.*, 2002; Bomberg et al 2015, see also **section 5.3.2**).

No obvious ferric iron reducing microbia were found to be amongst the most abundant taxa in BH10 (an *Acidiphilium*-related OTU was ranked 24 in abundance after OTUs occurring to the IC and DW controls were removed from the analysis, **Table 8.1**). Some members of the *Arthrobacter* genus (the most abundant actinobacterial taxa in BH10) have however been found capable of metal respiration (Ottow & Glathe, 1971; Jones *et al.*, 1984; Lee & Newman, 2003). Interestingly,

actinobacterial abundance was negatively correlated with iron and methane concentrations (**Figure 8.3**). On the other hand, the significant relationship between soluble ferric iron concentration upon the microbial community composition might be due to the presence of iron oxidizing microorganisms producing ferric iron as a product. However, if this were to be the case the potential iron oxidizer *Acidovorax* (*Betaproteobacteria*) would show a negative correlation with ferric iron concentration (**Figure 8.3**). These results are consistent with the lack of significant pyrite oxidation in BH10. In any case, let us note that the iron concentration values presented in this work should be interpreted with care, as only soluble iron was measured (see **section 4.4.2**). Both microbial iron oxidation and reduction can be performed by directly transferring electrons from or to undissolved mineral surfaces (Hernandez & Newman, 2001; Kostka *et al.*, 2002). Such processes are therefore often dependent upon the available mineral surface area rather than the dissolved mineral concentrations (Kostka *et al.*, 2002). Unfortunately, this variable was impossible to measure with the array of techniques available for this thesis.

Another potential acceptor for anaerobic methane oxidation is sulphate (Orphan *et al.*, 2001; Parkes *et al.*, 2007; Knittel & Boetius, 2009; Pedersen, 2013; Marlow *et al.*, 2014; see also **section 5.3.2**). *Deltaproteobacteria*, a class which hosts most of the known sulphate-reducing bacteria, was negatively correlated with methane concentration (**Figure 8.3**). However, as discussed above, the actual abundance of sulphate reducing bacteria in BH10 was practically negligible.

Finally, hydrogen concentration was negatively correlated with the abundance of the subsections II and III of the *Cyanobacteria* phylum. Metagenomic analysis of sample 420 revealed the presence of a versatile endolithic microbial community potentially supported by cyanobacterial hydrogenotrophy and the the degradation of refractory organic matter, probably similar to type III kerogen (see previous section). It might be possible that the cyanobacteria found in other samples are taking part in similar associations: samples 496 and 249, for example, shared with sample 420 the presence of subsections II and III of *Cyanobacteria*, as well as lower hydrogen concentrations than the rest of the samples (**Figure 8.1**; **Figure 8.2c**; **Figure 8.3**; **Annex 3 - Table S3**). On the other hand, the cyanobacteria present in the rest of the samples belonged to other taxa and their presence did not show a clear correlation with hydrogen concentration (**Figure 8.3**; **Annex 3 - Table S3**). In those cases, cyanobacteria might thus be playing a different role than the one described in the previous section.

The influence of heterotrophic metabolisms in the deep subsurface IPB ecosystem is hard to assess from our data. Secuences related to heterotrophs, including *T. lapidicaptus*, which was isolated from the nearby BH11, were also retrieved from BH10 samples (**section 8.2.4**). Metagenomic analysis

revealed the potential for kerogen-degrading activities in the microbial community of sample 420; however, those results can not be easily extended to the rest of the samples, given the heterogeneity of deep subsurface habitats. Organic carbon was present in all the samples tested, suggesting that kerogen might be a potential carbon and energy source throughout the borehole.

Organic acids, especially acetate, were detected in several samples (**Figure 8.1**). Their ultimate origin remains unclear, as they can be produced and consumed by many different microbial metabolisms. For example, in sample 420 (see previous section) carbon reserves (either extracellular EPS or internal reserves such as alphaproteobacterial PHBs) could be fermented to organic acids in the absence of oxygen or nitrate. Conversely, in the presence of electron acceptors acetate could be used to obtain energy via acetyl-CoA synthetase and tricarboxylic acid cycle, and used as a carbon source via the glyoxylate cycle. Therefore, sample 420 could be acting either as an acetate source or an acetate sink depending on electron acceptor availability, or acetate could be being cycled between fermentative and respiratory members of the community. This reasoning could also be extended to other metabolisms (e.g. acetogenesis versus acetoclastic methanogenesis). While small organic acids could be good carbon and electron shuttles, we find it difficult to believe that they are being exchanged between the different microbial communities found through the borehole, especially after considering the long distances involved, and the hydrogeological nature of the BH10 area (see below). We thus believe that the organic acids detected in the BH10 are the result of local fermentative metabolisms.

Overall, the BH10 subsurface has an extremely low permeability (Fernando Tornos, personal communication). This makes it unlikely that it is operating as a single ecological unit, as there is probably little opportunity for nutrient exchange between the different depths sampled in this study. This is supported by the lack of clear trends in community composition with depth, and also by the fact that two of the three most influential environmental parameters in shaping the microbial community were the concentrations of hydrogen and methane, two gases which can be generated abiotically (Hellevang *et al.*, 2011; Etiope & Lollar 2013; Okland *et al.*, 2014) and diffuse through rocks with relative ease. Nonetheless, even if most of the transect drilled in this study cuts through unfractured regions, the fact that fatty acid concentrations increased at the localized fracture zones suggests that water flow channels also exist in BH10, at least to a certain extent.

Access to pore-space and water might thus be a critical factor for surviving in compact continental deep subsurface ecosystems. Even although our samples were taken below the water table and that all non-isolated pores are expected to be flooded, the adsorption of water to silica surfaces might result in reduced water availability, especially in small pores with a high surface-to-volume ratio.

Fracture areas would provide water and soluble nutrients, but microorganisms might still need to bore through rocks in order to reach non-soluble compounds such as pyrite or kerogen. Our 16S results showed the presence of bacteria similar to those found in other oligotrophic, endolithic and arid environments (**section 8.2.4**). Those organisms, even if usually found in surface habitats such as deserts, tundra soils or permafrost, possess the necessary traits for colonizing deep subsurface habitats: they are adapted to inhabit rocks while operating under extremely low energy flows and withstanding severe environmental stresses.

Metagenomic analysis revealed the presence of a metabolically versatile endolithic microbial community in which several organisms cooperated in order to exploit the scarce resources available in the IPB subsurface (**Figure 8.6**, previous section). We believe that this included not only metabolic coupling, but also the accomplishment of other tasks that can be beneficial to the whole community, such as the production of an extracellular matrix that facilitates water and nutrient retention, or the active weathering of rocks in order to increase available space and gain access to novel sources of carbon and energy. Their association might therefore be an obligate one, with cooperation being the only way to fulfil all the requirements for surviving in the harsh continental deep subsurface.

9. OVERALL DISCUSSION

9.1. Novel insights in deep subsurface microbiology

9.1.1. Short-scale geological features determine microbial community composition in deep subsurface ecosystems

Several boreholes were performed in the Peña de Hierro area within the context of the MARTE and IPBSL projects (see **section 3**). Boreholes BH8 and BH10 were thoroughly studied using an array of complementary techniques in order to gain a deeper understanding of the biogeochemical processes occurring in the deep subsurface of the IPB. This led to the characterisation of two distinct underground habitats located in a unique geological setting which has several implications for the study of subsurface-surface interactions, early life environments or Mars habitability (see next sections).

The rocks in the BH8 borehole were fractured, and its oxidised waters were in equilibrium with the nearby Peña de Hierro mine pit (Fernando Tornos, personal communication). The co-occurrence of pyrite, water and oxygen leads to a microbial community focused on pyrite oxidation, which results in the production of high levels of sulphuric acid. Some of the sulphate is in turn used an electron acceptor in the anaerobic part of the borehole (**Figure 5.7**), but a large part of it probably remains on the aquifer, leading to groundwater acidification and the subsequent dissolution of ferric iron. BH8 is therefore an example of the underground biological reactor proposed by several authors (Fernández-Remolar *et al.*, 2008; 2008b; Amils *et al.*, 2013; 2014; Gómez-Ortiz *et al.*, 2014) in order to explain the unique conditions the Tinto River, which actually originates near the Peña de Hierro mine pit.

In spite of their proximity (c.a. 500 metres), and of the fact that they share a common geological setting, the BH10 borehole was very different from BH8. The BH10 area was generally non-fractured, and its aquifer hosted anoxic waters. The pyrite concentration was lower, and it was unoxidised. No evidences of pyrite oxidation or sulphate reduction were found below the water table. Instead, nitrate seemed to be the primary electron acceptor, and hydrogen, methane or buried kerogen were the most likely electron donors (**section 8.3.3**). While the BH8 microbial community was to a large extent aerobic and similar to that of an acid mine drainage, the BH10 conditions appeared to select for microorganisms with strong adaptations towards endolithic lifestyles and the tolerance of severe nutritional and environmental stresses. Similarly, where the BH8 subsurface ecosystem could be divided into three main compartments showing a certain degree of nutrient

exchange (**Figure 5.7**), the microbial communities found in the different samples of the BH10 borehole could not be assembled together into a comprehensive model of the whole column (**section 8.3.3**). A detailed metagenomic analysis of one of these samples revealed a microbial consortium in which the partners established complex physiological and ecological interactions. However, according to our results this consortium relied on elements that were already buried (kerogen, pyrite surfaces), likely of an abiotic origin (hydrogen) or present in the whole Tinto-Odiel aquifer (nitrate), being thus independent from any other process potentially occurring in the rest of the borehole. We therefore concluded that, unlike BH8, the BH10 borehole is not operating as a single ecological unit (**section 8.3.3**), and that instead each sample probably contained its own, independent ecosystem. Metagenomic analyses of more of the samples is underway in order to further explore this possibility.

Our results highlight the key importance of the short-scale geological context in shaping deep subsurface microbial ecosystems, with an emphasis on the influence of water availability. In compact rocks with low-degrees of fracture, microorganisms might be living within enclosed microniches with a high degree of isolation. Physical and physicochemical characterisation of such microniches, and the way the interact between themselves and the main fault zones, is required in order to deepen the understanding on the actual short-scale environment that deep subsurface microorganisms live in, and to better put in context our metagenomic results.

9.2.2. Methodological developments

Deep subsurface drilling projects are complex and expensive, and the resulting samples are unique and can not be replaced. In addition to this, performing metagenomic analyses of deep subsurface samples is particularly challenging due to the very low biomass abundance and the undesired potential interactions between nucleic acids and the mineral matrices. The low yield of nucleic acids isolation techniques in deep subsurface samples makes their metagenomic analysis especially vulnerable to contamination, as even trace amounts of contaminants might overshadow the genuine environmental nucleic acids. This makes the adoption of suitable laboratory protocols a critical step for the retrieval of meaningful results.

On this thesis we have developed and validated an optimized sample processing pipeline for deep subsurface metagenomics, that includes an adapted DNA extraction and amplification protocol using sterile reagents, and the use of novel bioinformatic algorithms to reduce the impact of sequencing errors and assess the degree of contamination present in metagenomic libraries (**section**

7.3). Together, these protocols provides high-quality, contamination-free metagenomic data even from mineral matrices lacking in biomass, being of general interest not only for the study of deep subsurface environments, but for any other microbial ecology project that involves the analysis of challenging samples.

9.2.3. Novel hypotheses generated by this work

The analysis of the BH10 microbial populations revealed an unexpected abundance of cyanobacteria in the deep subsurface of the IPB. These cyanobacteria did not originate from contamination events, and instead must be playing an important role in their respective communities. After performing a metagenomic analysis of sample 420 (BH10), we were able to propose novel electron transport chains that would allow for the survival of cyanobacteria under dark anoxic conditions. Interestingly, those mechanisms rely on traits already documented in *Cyanobacteria*, such the derivation of hydrogen electrons to an electron transport chain, the existence of a quinol-dependent nitrous oxide reductase or the derivation of electrons to extracellular acceptors through cytochrome bd when cytochrome b6f is overreduced. However, and to the best of our knowledge, we are the first ones to have linked these processes together to propose operative dark, anaerobic cyanobacterial electron transport chains using hydrogen as the main electron donor. Our hypothesis is supported by the negative correlation found between hydrogen concentrations and the presence of the subsection III of *Cyanobacteria*, and for the presence of all the required genes in the cyanobacterial fraction of our metagenome.

In addition to this, our model (**Figure 8.6**) highlights a new potential role for pyrite in the anoxic IPB subsurface. Pyrite surface chemistry has been accounted for the degradation of organic pollutants, such as trichloroethene, under anoxic conditions (Weerasooriya & Dharmasena, 2001), but to the best of our knowledge its potential role as a provider of oxidizing power in deep subsurface ecosystems has not been completely studied. Davila *et al.*, (2008) proposed that hydroxyl radical formation by pyrite could be responsible for the degradation of organic matter in the IPB – and potentially also in Mars – subsurface. Our metagenomic results showed the presence of a rock-inhabiting organic matter-degrading consortium in the BH10 that would require either oxygen or hydrogen peroxide for performing the initial steps of kerogen breakdown. We propose that pyrite is therefore a key resource in the IPB subsurface, and that endolithic microorganisms might be able to locate it – following oxygen gradients – and actively bore through rocks in order to exploit it.

9.2. Implications of the BH8 and BH10 deep subsurface ecosystems

9.2.1. The Iberian Pyrite Belt and the Tinto River Origin

The striking differences between the two boreholes can be explained by the different degree of fracture in both areas. Fractures bring in water and oxygen from upper regions of the borehole, but also facilitate the exchange of nutrients between different points, effectively homogenizing the otherwise heterogeneous continental subsurface, and allowing for the generation of biogeochemical gradients. In this sense, it is likely that the proximity of BH8 to the mine pit is at least partially responsible for the deep subsurface ecosystem observed in that area. Thus, while our results prove beyond any reasonable doubt that the extreme conditions in the Tinto River are strongly conditioned by biological processes, the comparison between BH8 and BH10 showed that, while a subsurface bioreactor does exist in the Peña de Hierro area, active underground pyrite oxidation is not necessarily a general feature of the IPB subsurface. It is unclear whether favourable conditions for extensive deep subsurface pyrite oxidation existed in the Peña de Hierro area before the starting of mining activity. Fernandez-Remolar *et al.* (2003; 2005) found evidences of massive laminated iron bioformations in the nearby Alto de la Mesa, which dated from 2.1 million years ago. This showed the existence of an ancient, non-anthropic iron cycle in the Tinto River origin, however at the present time the acidic conditions in the Tinto River are not limited to its headwaters, but extend through its whole course. According to a review from Olías & Nieto (2012), there is historical evidence of a significant deterioration of water quality downstream of Cerro Salomón (located two kilometres south of Peña de Hierro) during the second half of the nineteenth century. We therefore believe that, even if the unique conditions found in the Tinto River origin are a natural, non-anthropic phenomenon, its actual scope and impact on downstream waters have been exacerbated due to mining activity.

9.2.2. The Iberian Pyrite Belt subsurface as an early biosphere analogue

Prior to the rise of oxygen, the Archaean biosphere is believed to have been dominated by iron chemistry (Gómez & Amils, 2002; Amils *et al.*, 2005), due to the high availability of iron before the Great Oxygenation Event. Additionally, subsurface habitats have been proposed to be favourable shelters for early life, providing both nutrients and protection from radiation (Trevors, 2002). Our results showed the presence of an operative iron-sulphur cycle in the deep subsurface of the IPB

(**section 5.3.2**). Under the present conditions this system relies on oxygen, which was not present in the Archaean, for the initial oxidation of pyrite. However, and as described by Davila *et al.*, (2008), the reaction of pyrite with water is able to generate oxidants under dark, anoxic conditions, thus potentially pumping the iron and sulphur cycles. We believe that this is occurring in the less fractured BH10, albeit the resulting oxidants are probably being scavenged by microorganisms and used to degrade refractory organic matter. This reaction has also been hypothesized to be responsible for the early evolution of photosynthetic organisms, as the peroxidases and catalases required for hydrogen peroxide damage are structural analogues of photosynthetic reaction centres (Borda *et al.*, 2001). Thus, even although the IPB subsurface is currently inhabited by modern microorganisms, its overall conditions and physicochemistry is probably similar to that of ancient environments, making it a good proxy to test novel hypotheses on the origin and early development of life.

9.2.3. The Iberian Pyrite Belt subsurface as a Mars analogue

The Tinto River and its surrounding area are perhaps most known for their analogies to Martian environments, which have resulted in a large body of work (see Amils *et al.*, 2014 for a recent review). It has been proposed that a hypothetical Martian subsurface biosphere relying on chemolithotrophic metabolisms would be able to maintain pH and thermal homoeostasis, resulting in a more habitable environment (Fernández-Remolar *et al.*, 2008b). The iron and sulphur cycle-based community found in the BH8 would be an example of this. As discussed above, the reaction of pyrite with water would be able to generate oxidants even under dark, anoxic conditions (Davila *et al.*, 2008). Therefore, similar processes such as the ones described in this thesis for BH8 might have been operative under wet conditions such as the ones proposed for Mars paleoclimate (Grotzinger *et al.*, 2015). However, modern Martian environments are most likely water-limited, and our results in BH10 suggest that an ecosystem based on pyrite oxidation might not be possible without an extensive input of water (see above). Instead, we found evidences of an ecosystem based on hydrogenotrophy and ancient organic matter degradation in the BH10 borehole. Hydrogen is speculated to be available in significant quantities in modern Martian subsurface habitats, which the lack of water availability being the key factor limiting biological activity (Weiss *et al.*, 2000). This is similar to what we propose for the unfractured BH10 area. We therefore believe that the two boreholes studied in this thesis might be good analogues of two different Martian environments: an early, wet, habitable subsurface in which the matter and energy fluxes might have been high enough

for the development of a homoeostatic biosphere (BH8 borehole), and a modern, desiccated subsurface in which even the most resistant microorganisms would only be able to survive at very low metabolic speeds (BH10 borehole).

9.3. Future perspectives

The results presented in this thesis allowed us to find biological evidence of the deep subsurface origin of the Tinto River unique conditions, and to limit this phenomenon to the area closest to the Peña de Hierro mine pit. Additionally, the metagenomic analysis of the BH10 samples allowed us to characterise a novel deep subsurface microbial consortia, and to generate hypotheses regarding the potential roles of cyanobacteria in dark, deep subsurface environments. From the geobiological point of view, the IPB subsurface proved to be heterogeneous, with striking differences between BH8 and BH10 in both their chemical and microbiological composition. We concluded that the BH10 was inhabited of independent microbial communities occupying isolated microniches, but with the array of techniques used in this thesis we were unable to determine whether they were completely isolated or the internal pore structure of the BH10 rocks allowed for a biologically relevant exchange of nutrients between the different underground habitats.

It is worth noting that, even if we hypothesize that there is little-to-no interaction between the microbial communities found at the different depths of BH10, our reasoning only applies to that specific vertical transect, and does not preclude the existence of a complex ecosystem encompassing different zones of the Peña de Hierro deep subsurface. However, in order to study this, it would be first required to perform a comprehensive hydrogeological study in order to locate the major groundwater flow channels in the area. Only then could those channels be intersected at different locations and depths by vertical drillings, allowing the study of the different microbial communities present across the water flow path. A resistivity study of the area was performed and published by Gómez-Ortiz *et al.* (2014); however the selected transect did not intercept any of the boreholes studied in this thesis (see Gómez-Ortiz *et al.* 2014, Figure 1). The fact that intact fatty acids, used here as a proxy for extant or recent biomass, increased in the most fractured areas of BH10 does indeed suggest the existence of transverse faults carrying meteoric water; a hypothesis that was proposed by Gómez-Ortiz *et al.* (2014). The results presented in this thesis suggest that the deep IPB subsurface ecosystem is much more complex than anticipated, as very different biogeochemical processes can be operating within very short geographical distances. Even if additional drillings are not performed due to their considerable cost in time and money, more

detailed knowledge about the deep subsurface water flow in the Peña de Hierro area would surely be useful for building a model that integrates the results obtained for both BH8 and BH10.

Due to the heterogeneous nature of the continental subsurface, descriptive ecological studies that rely only on bulk mineralogy information – such as the one performed in this thesis – will likely not be able to provide a complete picture of the processes occurring in such environments. In this sense, and as stated before, the BH8 borehole was an exception, as its high degree of fracture reduced heterogeneity and the microbiological results could be easily fitted into an already well-studied paradigm (e.g. biological pyrite oxidation). However, interpreting the BH10 borehole results proved to be a far bigger challenge, and we were in fact unable to integrate the results found at the different depths into a single comprehensive model of the ecosystem. We have hypothesized about the potential challenges posed to microorganisms in the BH10 borehole, namely a lack of water, nutrients and space. This is consistent with the reports of a highly compact mineral matrix in that area (Fernando Tornos, personal communication). However, an actual study of the permeability, pore water chemistry (in opposition to results obtained using bulk samples) and pore size and distribution of the BH10 samples is still missing. We believe such a study would greatly help to characterise the actual microniches available for colonization in the BH10 subsurface, and to better define the specific conditions that the microbial communities found in this thesis are living on. Furthermore, identifying the main groundwater flow paths would help in integrating the results found for the different areas and depths into a single model.

Obtaining a general knowledge of deep subsurface biological processes is of course not an easy task. In this thesis we have focused on comparing two boreholes drilled in the Peña de Hierro area. By obtaining first-hand samples from this unique environment, we were able it to provide a preliminary characterisation of its biology, thus confirming hypotheses such as the existence of the underground Peña de Hierro bioreactor. We also took advantage of its analogies to potential early and extant Martian subsurface habitats, describing microbial communities that could be operative under such extreme conditions. However, the high cost of deep subsurface drilling projects results in a lack of samples and replication – a vertical transect will provide a descriptive one-time snapshot of the studied system. Answering complex questions such as the effect of different mineral matrices (including the role of pyrite), water regimes or pore water chemistries in shaping deep subsurface microbial communities requires repeated measurements under a wide array of conditions, something that can not be realistically attained in deep subsurface drilling projects. Therefore, we believe that research should also focus in the development of simpler model systems in which matter and energy fluxes can be monitored in a reproducible way. Microbial colonization

of rocks, and the effect of different pore structures and pore water chemistries could be simulated by the use of synthetic mineral matrices. Microfluidic systems could in turn prove useful to test the effect of different micro-scale water regimes on microorganism growth. Such experiments would require the use of culturable deep subsurface microbial strains. The *Tessaracoccus lapidicaptus* strain described in this thesis could be a good starting point for such experiments. Further work should focus in the characterisation of its genome and its potential for the degradation of refractory organic matter. Our results also suggested that some *Cyanobacteria* might be especially well adapted for thriving in deep subsurface conditions. Based on the metagenomic data, we have proposed two different lithoautotrophic energy transduction pathways for the cyanobacteria found in sample 420 (BH10). Confirming these dark, anoxic metabolic capabilities in a culturable cyanobacterium with similar genetic traits would have tremendous implications, and would provide researchers with an invaluable model species for studying endolithic and deep subsurface lifestyles.

10. CONCLUSIONS

- Geomicrobiological comparison of two deep drills in the IPB

The microbiology and physicochemistry of two deep drills (BH8 and BH10) were characterised and compared in order to gain understanding of the biogeochemical processes operating in the deep subsurface of the IPB.

- Improved processing of deep subsurface samples

We have developed and validated an integral “from-dirt-to-data” deep subsurface sample processing pipeline that covers sterile sample retrieval, sterile nucleic acids extraction and amplification from low biomass mineral samples, and improved bioinformatic processing of high throughput sequences. This pipeline allows us to minimize the possibility of sample contamination during sample processing, and to assess the degree of contamination remaining in the sequenced datasets.

- A biological origin for the Tinto River

The BH8 borehole area is generally fractured, possibly due to its closeness to the Peña de Hierro mine pit. Due to the input of oxygenated water to the massive pyrite stockwork, the BH8 ecosystem contained fully operative iron and sulphur cycles driven by pyrite oxidizing microorganisms. This corresponds, in both location and nature, to the underground bioreactor proposed by many authors as the responsible for the unique conditions of the Tinto River. Our results prove beyond any reasonable doubt that the Tinto River conditions have a strong underground biological component.

- Water availability is a key factor in deep subsurface environments

The BH10 area is less fractured, and its underground pyrite is mostly non-oxidised. Pyrite oxidizing microorganisms are not present in BH10 in significant amounts. Instead, the ecosystem is dominated by polyextremotolerant microorganisms. Resistance to desiccation appears to be a common trait of the species found in BH10, which suggests that under limiting pore sizes matric water stress might be high even when below the water table. In this context, the presence of cracks will result in an increase in the water influx, as well as in the space available for microbial colonization.

- Underground pyrite oxidation is not a general feature of the IPB subsurface

The differences between BH8 and BH10 indicate that extensive underground pyrite oxidation requires of a relatively high water input, and is not necessarily a general feature of the IPB subsurface. We therefore believe that, even if the unique conditions found in the Tinto River origin

are a natural, non-anthropic phenomenon, their actual scope and impact on downstream waters have been exacerbated due to the exposure of minerals to oxic waters as a consequence of mining activity.

- Diverse Mars analogue habitats

BH8 and BH10 also can be considered as analogues of two different stages in Martian subsurface evolution: an early, wet, habitable subsurface in which the matter and energy fluxes might have been high enough for the development of a homoeostatic biosphere (BH8 borehole), and a more recent, desiccated subsurface in which even the most resistant microorganisms would only be able to survive at very low metabolic rates (BH10 borehole).

- A novel deep subsurface microbial species

Enrichment cultures from deep subsurface samples allowed us to isolate and characterise a novel species from the genus *Tessaracoccus*, which we named *Tessaracoccus lapidicaptus* (la.pi.di.cap'tus. L. n. *lapis lapidis* stone; L. adj *captus* captive; N.L. masc. adj. *Lapidicaptus* stone-captive). Our experiments showed that *T. lapidicaptus* is able to drive the biotic precipitation of iron carbonates and phosphates. We also propose that it is adept at degrading refractory organic matter using nitrate as an electron acceptor, and thus plays a significant role in the IPB subsurface ecosystem.

- The metagenome of a novel dark subsurface cyanobacterium

Many of the BH10 samples were dominated by cyanobacteria similar to those found in arid and oligotrophic environments such as endoliths, deserts, tundras or permafrost. These results are not due to contamination during core sampling and processing and instead correspond to endogenous underground organisms. Statistical analysis revealed that the presence of the subsection III of *Cyanobacteria* was significantly correlated with a decrease in hydrogen concentrations. Our metagenomic analysis suggested that they could have a functional dark, anaerobic electron transport chain that couples hydrogen oxidation to the reduction of either nitrous oxide or extracellular acceptors. This hypothesis relies on traits that have been already documented in *Cyanobacteria*, but to the best of our knowledge have not been yet integrated together to propose a functional electron transport chain. While our hypothesis requires both a thermodynamic and an experimental validation, the unequivocal predominance of *Cyanobacteria* in many of our deep subsurface samples calls for a re-evaluation of the potential roles of these organisms in dark, anoxic environments.

- A versatile endolithic microbial consortium

These cyanobacteria formed part of a versatile endolithic microbial community, which was also composed by members from the *Ascomycota*, *Alphaproteobacteria* and *Bacteroidetes* groups. The main energy sources of the consortium appeared to be hydrogen oxidation and the degradation of refractory organic matter. Pyrite might be playing a key role in this ecosystem by providing oxidants via surface chemistry, even in the absence of oxygen and light. We believe that the different partners in the consortium interacted not only via metabolic coupling but also by being specialized in other tasks that can be beneficial to the whole community, such as the production of an extracellular matrix that facilitates water and nutrient retention, or the active weathering of rocks in order to increase available space and gain access to novel sources of carbon and energy. Their association might therefore be an obligate one, with cooperation being the only way to fulfil all the requirements for surviving in the harsh continental deep subsurface.

11. CONCLUSIONES

- Comparación geomicrobiológica de dos perforaciones en la Faja Pirítica Ibérica (FPI)

La microbiología y fisicoquímica de dos perforaciones (BH8 y BH10) fueron caracterizadas y comparadas con el objetivo de comprender los procesos biogeoquímicos que operan en el subsuelo profundo de la FPI.

- Procesamiento optimizado de muestras de subsuelo profundo

Hemos desarrollado y validado un protocolo integral para el procesamiento de muestras de subsuelo profundo, que cubre la recuperación de muestras estériles, la extracción y amplificación de ácidos nucleicos a partir de muestras minerales con baja biomasa, y el procesamiento mejorado de datos provenientes de plataformas de secuenciación de alto rendimiento. Este procedimiento nos permite minimizar la posibilidad de contaminaciones durante el procesado de las muestras, así como evaluar *a posteriori* el grado de contaminación remanente en los datos finales.

- Un origen biológico para Río Tinto

La zona de la perforación BH8 presenta, en términos generales, un grado alto de fractura, posiblemente debido a su cercanía a la corta de Peña de Hierro. Debido a la entrada de aguas óxicas en el sistema de pirita masiva, el ecosistema encontrado en la perforación BH8 contenía ciclos del hierro y del azufre totalmente operativos, impulsados por micoorganismos oxidadores de la pirita. El ecosistema observado corresponde, tanto en naturaleza como en localización física, al bioreactor subterráneo propuesto por varios autores como el responsable de las condiciones únicas del Río Tinto. Nuestros resultados prueban que las condiciones encontradas en el Río Tinto se deben en gran medida a la actividad biológica subterránea.

- La disponibilidad de agua es un factor clave en ecosistemas de subsuelo profundo

La zona de la perforación BH10, por contra, se halla menos oxidada, y su pirita mayormente no presenta signos de oxidación. No se encontraron organismos oxidadores de pirita en cantidades significativas. En vez de esto, el ecosistema está dominado por microorganismos poliextremotolerantes. La resistencia a la desecación parece ser una característica común de las especies encontradas en la perforación BH10, lo que sugiere que, para tamaños de poro lo suficientemente bajos, el estrés mátrico puede ser un factor relevante incluso por debajo del nivel freático. En este contexto, la presencia de grietas resultará en una mayor entrada de agua, así como en una mayor disponibilidad de espacio colonizable.

- La oxidación subterránea de pirita no es una característica general del subsuelo de la FPI

Las diferencias entre las perforaciones BH8 y BH10 indican que la oxidación subterránea de pirita requiere de un aporte de agua relativamente alto, y no es necesariamente una característica general del subsuelo de la FPI. Creemos por tanto que, incluso aunque las condiciones únicas encontradas en el Río Tinto son de origen no antrópico, la extensión de este fenómeno y su impacto río abajo se han visto exacerbadas debido a la exposición de minerales a aguas óxicas como consecuencia de la actividad minera.

- Diversidad de hábitats análogos de Marte en el subsuelo de la FPI

Las perforaciones BH8 y BH10 también pueden ser consideradas como análogas de dos etapas distintas en la evolución del subsuelo marciano: un subsuelo temprano, húmedo y habitable en el que los flujos de materia y energía eran lo suficientemente altos como para que se desarrollase una biosfera homeostática (BH8), y un subsuelo más reciente y desecado en el que incluso los microorganismos más resistentes serían capaces de sobrevivir únicamente a tasas metabólicas muy bajas.

- Una nueva especie microbiana encontrada en el subsuelo profundo

El uso cultivos de enriquecimiento sobre muestras de subsuelo profundo nos permitió aislar y caracterizar una nueva especie del género *Tessaracoccus*, que bautizamos como *Tessaracoccus lapidicaptus* (la.pi.di.cap'tus. L. n. *lapis lapidis* piedra; L. adj *captus* cautiva; N.L. adj. masc. *Lapidicaptus* “cautiva en la piedra”). Nuestros experimentos mostraron que *T. lapidicaptus* es capaz de impulsar la precipitación biológica de carbonatos y fosfatos de hierro. También proponemos que es capaz de degradar materia orgánica recalcitrante usando nitrato como aceptor de electrones, y que por lo tanto cumple un rol importante en el ecosistema del subsuelo de la FPI.

- El metagenoma de una nueva cianobacteria subterránea

Muchas de las muestras de la perforación BH10 estaban dominadas por cianobacterias similares a aquellas encontradas en ambientes oligotróficos y áridos como endolitos, desiertos, tundras o permafrost. Estos resultados no se deben a contaminaciones durante la recuperación y el procesado de testigos, sino que corresponden a organismos nativos del subsuelo. Los análisis estadísticos revelaron que la presencia de la subsección III del phylum *Cyanobacteria* estaba significativamente correlacionada con el descenso de las concentraciones de hidrógeno. Nuestro análisis metagenómico sugiere que podrían tener una cadena de transporte de electrones funcional en ausencia de luz y oxígeno, acoplando la oxidación de hidrógeno a la reducción de óxido nitroso o

aceptores extracelulares. Esta hipótesis se sustenta en características que ya han sido documentadas previamente en el phylum *Cyanobacteria*, pero que, hasta donde sabemos, no han sido aún integradas para proponer una cadena de transporte de electrones funcional. Si bien esta hipótesis ha de ser validada tanto termodinámica como experimentalmente, la predominancia inequívoca de cianobacterias en muchas de nuestras muestras de subsuelo profundo llama a una reevaluación del papel de estos organismos en ambientes libres de luz y oxígeno.

- Un consorcio microbiano endolítico y metabólicamente versátil

Estas cianobacterias formaban parte de una comunidad endolítica metabólicamente versátil, que se hallaba también compuesta por miembros de los grupos *Ascomycota*, *Alphaproteobacteria* y *Bacteroidetes*. Las fuentes principales de energía de este consorcio son aparentemente la oxidación de hidrógeno y la degradación de materia orgánica recalcitrante. La pirita podría estar jugando un papel clave en este ecosistema, al proporcionar moléculas oxidantes gracias a reacciones químicas en su superficie, incluso en la ausencia de luz y oxígeno molecular. Creemos que los distintos miembros de esta comunidad interaccionan no sólo por medio de acoplamientos metabólicos, sino también mediante la especialización en distintas tareas que pueden ser beneficiosas para el conjunto de la comunidad, como la producción de una matriz extracelular que facilite la retención de agua y nutrientes, o el desgaste activo de las rocas colindantes con el objetivo de incrementar el espacio disponible y acceder a nuevas fuentes de carbono y energía. Su asociación podría por tanto ser obligada, siendo la cooperación entre distintos microorganismos la única manera de cumplir todos los requerimientos necesarios para la supervivencia en las duras condiciones del subsuelo profundo.

11. REFERENCES

Abbey, W., Salas, E., Bhartia, R., & Beegle, L. W. (2013). The Mojave Vadose Zone: A Subsurface Biosphere Analogue for Mars. *Astrobiology*, **13**(7), 637-646.

Abraham, P. E., Giannone, R. J., Xiong, W., & Hettich, R. L. (2014). Metaproteomics: extracting and mining proteome information to characterize metabolic activities in microbial communities. *Current Protocols in Bioinformatics*, **13.26**.

Abram, J. W., & Nedwell, D. B. (1978). Inhibition of methanogenesis by sulphate reducing bacteria competing for transferred hydrogen. *Archives of Microbiology*, **117**(1), 89-92.

Achtnich, C., Bak, F., & Conrad, R. (1995). Competition for electron donors among nitrate reducers, ferric iron reducers, sulfate reducers, and methanogens in anoxic paddy soil. *Biology and Fertility of Soils*, **19**(1), 65-72.

Aguilera, A., Manrubia, S. C., Gómez, F., Rodríguez, N., & Amils, R. (2006). Eukaryotic community distribution and its relationship to water physicochemical parameters in an extreme acidic environment, Rio Tinto (Southwestern Spain). *Applied and environmental microbiology*, **72**(8), 5325-5330.

Akob, D. M., & Küsel, K. (2011). Where microorganisms meet rocks in the Earth's Critical Zone. *Biogeosciences Discussions*, **8**(2), 2523-2562.

Aloisi, G., Gloter, A., Krüger, M., Wallmann, K., Guyot, F., & Zuddas, P. (2006). Nucleation of calcium carbonate on bacterial nanoglobules. *Geology*, **34(**12), 1017-1020.

Amann, R. I. (1995). In situ identification of micro-organisms by whole cell hybridization with rRNA-targeted nucleic acid probes. In *Molecular microbial ecology manual* (pp. 331-345). Springer Netherlands.

Amaral-Zettler, L. A., Gómez, F., Zettler, E., Keenan, B. G., Amils, R., & Sogin, M. L. (2002). Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature*, **417**(6885), 137-137.

Amils, R., González-Toril, E., Gómez, F., Fernández-Remolar, D., Rodríguez, N., Malki, M., Aguilera, A., & Amaral-Zettler, L. A. (2005). Importance of chemolithotrophy for early life on earth: the Tinto River (Iberian Pyritic Belt) case. In Origins (pp. 463-480). Springer Netherlands.

Amils, R., Fernández-Remolar, D., Parro, V., Rodríguez-Manfredi, J. A., Timmis, K., Oggerin, M., Sánchez-Román, M., López, F. J., Fernández, J. P., Puente, F., Gómez-Ortiz, D., Briones, C., Gómez, F., Omoregie, E. O., García, M., Rodríguez, N., & Sanz, J. L. (2013). Iberian Pyrite Belt Subsurface Life (IPBSL), a drilling project of biohydrometallurgical interest. *Advanced Materials Research*, **825**, 15-18.

Amils, R., Fernández-Remolar, D., & IPBSL Team. (2014). Río Tinto: A Geochemical and Mineralogical Terrestrial Analogue of Mars. *Life*, **4**(3), 511-534.

Amos, R. T., Bekins, B. A., Cozzarelli, I. M., Voytek, M. A., Kirshtein, J. D., Jones, E. J. P., &

Blowes, D. W. (2012). Evidence for iron-mediated anaerobic methane oxidation in a crude oil-contaminated aquifer. *Geobiology*, **10**(6), 506-517.

Anbar, A. D., & Knoll, A. H. (2002). Proterozoic ocean chemistry and evolution: a bioinorganic bridge?. *Science*, **297**(5584), 1137-1142.

Ander, P., Eriksson, M. E., & Eriksson, K. E. (1985). Methanol production from lignin-related substances by Phanerochaete chrysosporium. *Physiologia Plantarum*, **65**(3), 317-321.

Ander, P., & Eriksson, K. E. (1985). Methanol formation during lignin degradation by Phanerochaete chrysosporium. *Applied microbiology and biotechnology*, **21**(1-2), 96-102.

Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**(6), 683-693.

Andrews, S. C. (1998). Iron storage in bacteria. *Advances in microbial physiology*, **40**, 281-351.

Andrews, S. C. (2010). The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica et Biophysica Acta (BBA)-General Subjects*, **1800**(8), 691-705.

Ascaso, C., Wierzchos, J., Souza-Egipsy, V., de los Rí, A., & Rodrigues, J. D. (2002). In situ evaluation of the biodeteriorating action of microorganisms and the effects of biocides on carbonate rock of the Jeronimos Monastery (Lisbon). *International biodeterioration & biodegradation*, **49**(1), 1-12.

Bach, W., & Edwards, K. J. (2003). Iron and sulfide oxidation within the basaltic ocean crust: implications for chemolithoautotrophic microbial biomass production. *Geochimica et Cosmochimica Acta*, **67**(20), 3871-3887.

Bach, W., & Früh-Green, G. L. (2010). Alteration of the oceanic lithosphere and implications for seafloor processes. *Elements*, **6**(3), 173-178.

Bach, W., Paulick, H., Garrido, C. J., Ildefonse, B., Meurer, W. P., & Humphris, S. E. (2006). Unraveling the sequence of serpentinization reactions: petrography, mineral chemistry, and petrophysics of serpentinites from MAR 15 N (ODP Leg 209, Site 1274). *Geophysical Research Letters*, **33**(13), 1-4.

Bae, H. S., Moe, W. M., Yan, J., Tiago, I., Da Costa, M. S., & Rainey, F. A. (2006). *Brooklawnia cerclae* gen. nov., sp. nov., a propionate-forming bacterium isolated from chlorosolvent-contaminated groundwater. *International journal of systematic and evolutionary microbiology*, **56**(8), 1977-1983.

Bagchi, S. N., Chauhan, V. S., & Palod, A. (1990). Heterotrophy and nitrate metabolism in a cyanobacteriumPhormidium uncinatum. *Current Microbiology*, **21**(1), 53-57.

Balkwill, D. L., Leach, F. R., Wilson, J. T., McNabb, J. F., & White, D. C. (1988). Equivalence of microbial biomass measures based on membrane lipid and cell wall components, adenosine triphosphate, and direct counts in subsurface aquifer sediments. *Microbial Ecology*,

16(1), 73-84.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A, D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455-477.

Baqué, M., Viaggiu, E., Scalzi, G., & Billi, D. (2013). Endurance of the endolithic desert cyanobacterium Chroococcidiopsis under UVC radiation. *Extremophiles*, **17**(1), 161-169.

Barberes, G. (2015). Organic geochemical results from the shales of the South Portuguese Zone, southern Portugal. In *European Regional Conference and Exhibition*.

Barbieri, E., Guidi, C., Bertaux, J., Frey-Klett, P., Garbaye, J., Ceccaroli, P., Saltarelli, R., Zambonelli, A., & Stocchi, V. (2007). Occurrence and diversity of bacterial communities in Tuber magnatum during truffle maturation. *Environmental Microbiology*, **9**(9), 2234-2246.

Barker, W. W., & Banfield, J. F. (1996). Biologically versus inorganically mediated weathering reactions: relationships between minerals and extracellular microbial polymers in lithobiontic communities. *Chemical Geology*, **132**(1), 55-69.

Barker, W. W., Welch, S. A., Chu, S., & Banfield, J. F. (1998). Experimental observations of the effects of bacteria on aluminosilicate weathering. *American Mineralogist*, **83**, 1551-1563.

Barragán, M. J. L., Carmona, M., Zamarro, M. T., Thiele, B., Boll, M., Fuchs, G., García, J. L., & Díaz, E. (2004). The bzd gene cluster, coding for anaerobic benzoate catabolism, in Azoarcus sp. strain CIB. *Journal of bacteriology*, **186**(17), 5762-5774.

Barsanti, L., & Gualtieri, P. (2014). *Algae: anatomy, biochemistry, and biotechnology*. CRC press. p50.

Bartosch, S., Wolgast, I., Spieck, E., & Bock, E. (1999). Identification of nitrite-oxidizing bacteria with monoclonal antibodies recognizing the nitrite oxidoreductase. *Applied and environmental microbiology*, **65**(9), 4126-4133.

Beeman, R. E., & Suflita, J. M. (1989). Evaluation of deep subsurface sampling procedures using serendipitous microbial contaminants as tracer organisms. *Geomicrobiology Journal*, **7**(4), 223-233.

Bernhard, M., Benelli, B., Hochkoeppler, A., Zannoni, D., & Friedrich, B. (1997). Functional and structural role of the cytochrome b subunit of the membrane-bound hydrogenase complex of Alcaligenes eutrophus H16. *European Journal of Biochemistry*, **248**(1), 179-186.

Bernal, J. D. (1949). The physical basis of life. *Proceedings of the Physical Society. Section B*, **62**(10), 597.

Biddle, J. F., Lipp, J. S., Lever, M. A., Lloyd, K. G., Sørensen, K. B., Anderson, R., Fredricks, H. F., Elvert, M., Kelly, T.J., Schrag, D.P., Sogin, M.L., Brenchley, J.E., Teske, A., House, C.H. & Hinrichs, K. U. (2006). Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(10),

3846-3851.

Bigham, J. M., Schwertmann, U., Traina, S. J., Winland, R. L., & Wolf, M. (1996). Schwertmannite and the chemical modeling of iron in acid sulfate waters. *Geochimica et Cosmochimica Acta*, **60**(12), 2111-2121.

Billi, D., Wright, D. J., Helm, R. F., Prickett, T., Potts, M., & Crowe, J. H. (2000). Engineering desiccation tolerance in *Escherichia coli*. *Applied and environmental microbiology*, **66**(4), 1680-1684.

Billi, D., & Potts, M. (2002). Life and death of dried prokaryotes. *Research in microbiology*, **153**(1), 7-12.

Blackhurst, R. L., Genge, M. J., Kearsley, A. T., & Grady, M. M. (2005). Cryptoendolithic alteration of Antarctic sandstones: pioneers or opportunists?. *Journal of Geophysical Research: Planets (1991–2012)*, **110**(E12).

Blake, R., & Johnson, D. B. (2000). Phylogenetic and biochemical diversity among acidophilic bacteria that respire on iron. In: D.R. Lovely (ed.) Environmental microbe-metal interactions, ASM Press, Washington, pp. 53- 78.

Blanco, Y., Prieto-Ballesteros, O., Gómez, M. J., Moreno-Paz, M., García-Villadangos, M., Rodríguez-Manfredi, J. A., Cruz-Gil, P., Sánchez-Román, M., Rivas, L. A., & Parro, V. (2012). Prokaryotic communities and operating metabolisms in the surface and the permafrost of Deception Island (Antarctica). *Environmental microbiology*, **14**(9), 2495-2510.

Blanco, Y., Rivas, L. A., García-Moyano, A., Aguirre, J., Cruz-Gil, P., Palacín, A., van Heerden, E., & Parro, V. (2014). Deciphering the Prokaryotic Community and Metabolisms in South African Deep-Mine Biofilms through Antibody Microarrays and Graph Theory. *PloS one*, **9**(12), e114180.

Bohorquez, L. C., Delgado-Serrano, L., López, G., Osorio-Forero, C., Klepac-Ceraj, V., Kolter, R., Junka, H., Baena, S., & Zambrano, M. M. (2012). In-depth characterisation via complementing culture-independent approaches of the microbial community in an acidic hot spring of the Colombian Andes. *Microbial ecology*, **63**(1), 103-115.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, **10**(1), 57-59.

Bolin B & Cook RB (1983) in: *The Major Biogeochemical Cycles and their Interactions*, John Wiley and Sons, New York.

Bollmann, A., Palumbo, A. V., Lewis, K., & Epstein, S. S. (2010). Isolation and physiology of bacteria from contaminated subsurface sediments. *Applied and environmental microbiology*, **76**(22), 7413-7419.

Bomberg, M., Nyyssönen, M., Pitkänen, P., Lehtinen, A., & Itävaara, M. (2015). Active Microbial Communities Inhabit Sulphate-Methane Interphase in Deep Bedrock Fracture Fluids in

Olkiluoto, Finland. *BioMed Research International*, 2015.

Bonaccorsi, R., & Stoker, C. R. (2008). Science results from a Mars drilling simulation (Río Tinto, Spain) and ground truth for remote science observations. *Astrobiology*, **8**(5), 967-985.

Bonte M, Röling WF, Zaura E, van der Wielen PW, Stuyfzand PJ & van Breukelen BM (2013). Impacts of shallow geothermal energy production on redox processes and microbial communities. *Environ. Sci. Technol*. **47**:14476-14484.

Bontognali, T. R., Vasconcelos, C., Warthmann, R. J., Dupraz, C., Bernasconi, S. M., & McKenzie, J. A. (2008). Microbes produce nanobacteria-like structures, avoiding cell entombment. *Geology*, **36**(8), 663-666.

Borda, M. J., Elsetinow, A. R., Schoonen, M. A., & Strongin, D. R. (2001). Pyrite-induced hydrogen peroxide formation as a driving force in the evolution of photosynthetic organisms on an early Earth. *Astrobiology*, **1**(3), 283-288.

Borda, M. J., Elsetinow, A. R., Strongin, D. R., & Schoonen, M. A. (2003). A mechanism for the production of hydroxyl radical at surface defect sites on pyrite. *Geochimica et Cosmochimica Acta*, **67**(5), 935-939.

Bosak, T., & Newman, D. K. (2003). Microbial nucleation of calcium carbonate in the Precambrian. *Geology*, **31**(7), 577-580.

Boston, P. J., Ivanov, M. V., & P McKay, C. (1992). On the possibility of chemosynthetic ecosystems in subsurface habitats on Mars. *Icarus*, **95**(2), 300-308.

Bothe, H., Schmitz, O., Yates, M. G., & Newton, W. E. (2011). Nitrogenases and hydrogenases in cyanobacteria. In *Bioenergetic Processes of Cyanobacteria* (pp. 137-157). Springer Netherlands.

Bouropoulos, N. C., & Koutsoukos, P. G. (2000). Spontaneous precipitation of struvite from aqueous solutions. *Journal of Crystal Growth*, **213**(3), 381-388.

Bradley, J. P., Harvey, R. P., McSween, H. Y., Gibson, E., Thomas-Keprta, K., & Vali, H. (1997). No ‘nanofossils’ in martian meteorite. *Nature*, **390**(6659), 454-456.

Brain, D. A., & Jakosky, B. M. (1998). Atmospheric loss since the onset of the Martian geologic record: Combined role of impact erosion and sputtering. *Journal of Geophysical Research: Planets* (1991–2012), **103**(E10), 22689-22694.

Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences*, **100**(7), 3960-3964.

Brinton, K. L., Tsapin, A. I., Gilichinsky, D., & McDonald, G. D. (2002). Aspartic acid racemization and age-depth relationships for organic carbon in Siberian permafrost. *Astrobiology*, **2**(1), 77-82.

Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W. L., Russ, C.,

Lander, E. S., Nusbaum, C., & Jaffe, D. B. (2008). Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome research*, **18**(5), 763-770.

Brown, A. D. (1976). Microbial water stress. *Bacteriological Reviews*, **40**(4), 803.

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., & Brom, T. H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv preprint arXiv:1203.4802*.

Bruce, C., Stone, K., Gulcicek, E. & Williams, K. (2013). Proteomics and the Analysis of Proteomic Data: 2013 Overview of Current Protein-Profiling Technologies. *Current Protocols in Bioinformatics*, **13.21**.

Brum, J. R., & Sullivan, M. B. (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology*.

Bruno, L., Billi, D., Bellezza, S., & Albertano, P. (2009). Cytomorphological and genetic characterisation of troglobitic Leptolyngbya strains isolated from Roman hypogea. *Applied and environmental microbiology*, **75**(3), 608-617.

Bugg, T. D., Ahmad, M., Hardiman, E. M., & Rahmanpour, R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Natural product reports*, **28**(12), 1883-1896.

Burdige, D. J. (2006). Geochemistry of marine sediments (Vol. 398). Princeton: Princeton University Press.

Butler, K. and Stephens, M. (1993). The distribution of a sum of binomial random variables. *Technology Report* STANFORD UNIV CA DEPT OF STATISTICS **467**, Stanford, USA.

Büsch, A., Friedrich, B., & Cramm, R. (2002). characterisation of the norB gene, encoding nitric oxide reductase, in the nondenitrifying cyanobacterium Synechocystis sp. strain PCC6803. *Applied and environmental microbiology*, **68**(2), 668-672.

Cai, M., Wang, L., Cai, H., Li, Y., Wang, Y. N., Tang, Y. Q., & Wu, X. L. (2011). *Salinarimonas ramus* sp. nov. and *Tessaracoccus oleiagri* sp. nov., isolated from a crude oil-contaminated saline soil. *International journal of systematic and evolutionary microbiology*, **61**(8), 1767-1775.

Camarinha-Silva, A., Jáuregui, R., Chaves-Moreno, D., Oxley, A., Schaumburg, F., Becker, K., Wos-Oxley, M. L., & Pieper, D. H. (2014). Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. *Environmental microbiology*, **16**(9), 2939-2952.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N, González-Peña, A., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, **7**(5), 335-336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N. and Knight, R. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, **108**(Supplement 1), 4516-4522.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, **6**(8), 1621-1624.

Carvalho, D., Barriga, F. J. A. S., & Munha, J. (1999). Bimodal siliciclastic systems - the case of the Iberian Pyrite Belt. In: Barrie, C.T. and Hannington, M. D. (Eds.), Volcanic-associated Massive Sulfide Deposits: Processes and Examples in Modern and Ancient Settings, Reviews in Economic Geology, vol. 8 , SEG, p. 375-408.

Cashion, P., Holder-Franklin, M. A., McCully, J., & Franklin, M. (1977). A rapid method for the base ratio determination of bacterial DNA. *Analytical biochemistry*, **81**(2), 461-466.

Cavaller, J. R. A., Robles, R. C., Rebollo, A. A., & Otero, R. V. (2005). Caracterización cuantitativa y cualitativa de las masas de agua subterránea en las unidades hidrológicas en el ámbito del Plan Hidrológico del Guadiana II (cuencas de los ríos Tinto, Odiel y Guadiana). In *VI Simposio del Agua en Andalucía: 1 a 3 de junio, 2005 Sevilla* (pp. 407-416).

Chapelle, F. H., & McMahon, P. B. (1991). Geochemistry of dissolved inorganic carbon in a Coastal Plain aquifer. 1. Sulfate from confining beds as an oxidant in microbial CO_2 production. *Journal of Hydrology*, **127**(1), 85-108.

Chapelle, F. H., O'Neill, K., Bradley, P. M., Methé, B. A., Ciufo, S. A., Knobel, L. L., & Lovley, D. R. (2002). A hydrogen-based subsurface microbial community dominated by methanogens. *Nature*, **415**(6869), 312-315.

Chilingarian, G. V. (1983). Compactional diagenesis. In *Sediment diagenesis* (pp. 57-168). Springer Netherlands.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L.-H., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L.M., Andersen, G. M., Hazen, T. C., Brockman, F. J., Arkin, A. P., & Onstott, T. C. (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science*, **322**(5899), 275-278.

Clegg, J. S. (2001). Cryptobiosis—a peculiar state of biological organization. Comparative Biochemistry and Physiology Part B: *Biochemistry and Molecular Biology*, **128**(4), 613-624.

Cody, G. D., Boctor, N. Z., Filley, T. R., Hazen, R. M., Scott, J. H., Sharma, A., & Yoder, H. S. (2000). Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate. *Science*, **289**(5483), 1337-1340.

Cohn, C. A., Mueller, S., Wimmer, E., Leifer, N., Greenbaum, S., Strongin, D. R., & Schoonen, M. A. (2006). Pyrite-induced hydroxyl radical formation and its effect on nucleic acids. *Geochemical Transactions*, **7**(3), 1-11.

Cole, J. K., Hutchison, J. R., Renslow, R. S., Kim, Y. M., Chrisler, W. B., Engelmann, H. E., Dohnalkova, A. C., Hu, D., Metz, T. O., Fredrickson, J. K., & Lindemann, S. R. (2014). Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Frontiers in microbiology*, **5**.

Colwell, F. S., Onstott, T. C., Delwiche, M. E., Chandler, D., Fredrickson, J. K., Yao, Q. J., McKinley, J. P., Boone, D. R., Griffiths, R., Phelps, T. J., Ringelberg, D., White, D. C., LaFreniere, L., Balkwill, D., Lehman, R. M., Konisky, J., & Long, P. E. (1997). Microorganisms from deep, high temperature sandstones: constraints on microbial colonization. *FEMS Microbiology Reviews*, **20**(3-4), 425-435.

Colwell, FS & RP Smith (2004). "Unifying principles of the deep terrestrial and deep marine biospheres." In *The Subseafloor Biosphere at Mid-ocean Ridges*, eds. Wilcock WSD, DeLong EF, Kelley DS, Baross JA & Cary SC (American Geophysical Union, Washington DC), pp. 355–367.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, **326**(5960), 1694-1697.

Cowen, J. P., Giovannoni, S. J., Kenig, F., Johnson, H. P., Butterfield, D., Rappé, M. S., Hutnak, M., & Lam, P. (2003). Fluids from aging ocean crust that support microbial life. *Science*, **299**(5603), 120-123.

Cragg, B. A., Parkes, R. J., Fry, J. C., Herbert, R. A., Wimpenny, J. W. T., & Getliff, J. M. (1990). *Bacterial biomass and activity profiles within deep sediment layers*. In Suess, E., von Huene, R., et al., Proc. ODP, Sci. Results (Vol. 112, pp. 607-619).

Crocker, F. H., Fredrickson, J. K., White, D. C., Ringelberg, D. B., & Balkwill, D. L. (2000). Phylogenetic and physiological diversity of Arthrobacter strains isolated from unconsolidated subsurface sediments. *Microbiology*, **146**(6), 1295-1310.

D'Hondt, S., Rutherford, S., & Spivack, A. J. (2002). Metabolic activity of subsurface life in deep-sea sediments. *Science*, **295**(5562), 2067-2070.

D'Hondt, S., Jørgensen, B. B., Miller, D. J., Batzke, A., Blake, R., Cragg, B. A., Cypionka, H., Dickens, G. R., Ferdelman, T., Hinrichs, K. U., Holm, N. G., Mitterer, R., Spivack, A., Wang, G., Bekins, B., Engelen, B., Ford, K., Gettemy, G., Rutherford, S. D., Sass, H., Skilbeck, C. G., Aiello, I. W., Guèrin, G., House, C. H,, Inagaki, F., Meister, P., Naehr, T., Niitsuma, S., Parkes, R. J,, Schippers, A., Smith, D. C., Teske, A., Wiegel, J., Padilla, C. N., & Acosta, J. L. (2004). Distributions of microbial activities in deep subseafloor sediments. *Science* **306**(5705), 2216-2221.

D'Hondt, S., Spivack, A. J., Pockalny, R., Ferdelman, T. G., Fischer, J. P., Kallmeyer, J., Abrams, L. J., Smith, D. C., Graham, D., Hasiuk, F., Schrum, H., & Stancin, A. M. (2009). Subseafloor sedimentary life in the South Pacific Gyre. *Proceedings of the National Academy of Sciences*, **106**(28), 11651-11656.

Daims, H., Brühl, A., Amann, R., Schleifer, K. H., & Wagner, M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and applied microbiology*, **22**(3), 434-444.

Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Venkateswaran, A., Hess, M., Kostandarithes, H. M., Makarova, K. S., Wacket, L. P., Fredrikson, J. K., & Ghosal, D. (2004). Accumulation of Mn (II) in Deinococcus radiodurans facilitates gamma-radiation resistance. *Science*, **306**(5698), 1025-1028.

Dash, J. G., Rempel, A. W., & Wettlaufer, J. S. (2006). The physics of premelted ice and its geophysical consequences. *Reviews of modern physics*, **78**(3), 695.

Davidson, M. M., Silver, B. J., Onstott, T. C., Moser, D. P., Gihring, T. M., Pratt, L. M., Boice, E. A., Lollar, B. S., Lippmann-Pipke, J., Pfiffner, S. M., Kieft, T. L., Seymore, W., & Ralston, C. (2011). Capture of planktonic microbial diversity in fractures by long-term monitoring of flowing boreholes, Evander Basin, South Africa. *Geomicrobiology Journal*, **28**(4), 275-300.

Davila, A. F., Fairén, A. G., Gago-Duport, L., Stoker, C., Amils, R., Bonaccorsi, R., Zavaleta, J., Lim, D., Schulze-Makuch, D., & McKay, C. P. (2008). Subsurface formation of oxidants on Mars and implications for the preservation of organic biosignatures. *Earth and Planetary Science Letters*, **272**(1), 456-463.

Davis Jr, R. A., Welty, A. T., Borrego, J., Morales, J. A., Pendon, J. G., & Ryan, J. G. (2000). Rio Tinto estuary (Spain): 5000 years of pollution. *Environmental Geology*, **39**(10), 1107-1116.

de Castro, A. P., Fernandes, G. D. R., & Franco, O. L. (2014). Insights into novel antimicrobial compounds and antibiotic resistance genes from soil metagenomes. *Frontiers in microbiology*, **5**.

de Duve, C., & Miller, S. L. (1991). Two-dimensional life?. *Proceedings of the National Academy of Sciences*, **88**(22), 10014-10017.

de los Ríos, A., Grube, M., Sancho, L. G., & Ascaso, C. (2007). Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS microbiology ecology*, **59**(2), 386-395.

de Macario, E. C., & Macario, A. J. (2000). Stressors, stress and survival: overview. *Frontiers in Biosciences*, **5**, D780-D786.

DeAngelis, K. M., Allgaier, M., Chavarria, Y., Fortney, J. L., Hugenholtz, P., Simmons, B., Sublette, K., Silver., W. L., & Hazen, T. C. (2011). characterisation of trapped lignin-degrading microbes in tropical forest soil. *PLoS One*, **6**(4), e19306-e19306.

DeAngelis, K. M., Sharma, D., Varney, R., Simmons, B., Isern, N. G., Markilllie, L. M., Nicora, C., Norbeck, A. D., Taylor, R. C., Aldrich, J. T., & Robinson, E. W. (2013). Evidence supporting dissimilatory and assimilatory lignin degradation in Enterobacter lignolyticus SCF1. *Frontiers in microbiology*, **4**.

Decho, A. W., Visscher, P. T., & Reid, R. P. (2005). Production and cycling of natural microbial exopolymers (EPS) within a marine stromatolite. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **219**(1), 71-86.

Delehanty, J. B., & Ligler, F. S. (2002). A microarray immunoassay for simultaneous

detection of proteins and bacteria. *Analytical Chemistry*, 74(21), 5681-5687.

Deming, J. W., & Baross, J. A. (1993). Deep-sea smokers: Windows to a subsurface biosphere?. *Geochimica et Cosmochimica Acta*, **57**(14), 3219-3230.

Ding, B., Schmeling, S., & Fuchs, G. (2008). Anaerobic metabolism of catechol by the denitrifying bacterium Thauera aromatica—A result of promiscuous enzymes and regulators?. *Journal of bacteriology*, **190**(5), 1620-1630.

Ding, T. & Schloss, P. (2014). Dynamics and associations of microbial community types across the human body. *Nature*, **509**(7500), 357-360.

Direito, S. O., Marees, A., & Röling, W. F. (2012). Sensitive life detection strategies for low-biomass environments: optimizing extraction of nucleic acids adsorbing to terrestrial and Mars analogue minerals. *FEMS microbiology ecology*, **81**(1), 111-123.

Dong, Y., Sanford, R. A., Locke, R. A., Cann, I. K., Mackie, R. I., & Fouke, B. W. (2014). Fe-oxide grain coatings support bacterial Fe-reducing metabolisms in 1.7−2.0 km-deep subsurface quartz arenite sandstone reservoirs of the Illinois Basin (USA). *Frontiers in microbiology*, **5**, 511.

Dose, K., Bieger-Dose, A., Labusch, M., & Gill, M. (1992). Survival in extreme dryness and DNA-single-strand breaks. *Advances in Space Research*, **12**(4), 221-229.

Dose, K., Bieger-Dose, A., Ernst, B., Feister, U., Gómez-Silva, B., Klein, A., Risi, S., & Stridde, C. (2001). Survival of microorganisms under the extreme conditions of the Atacama Desert. *Origins of Life and Evolution of the Biosphere*, **31**(3), 287-303.

Dupraz, C., Reid, R. P., Braissant, O., Decho, A. W., Norman, R. S., & Visscher, P. T. (2009). Processes of carbonate precipitation in modern microbial mats. *Earth-Science Reviews*, **96**(3), 141-162.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**(19), 2460-2461.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**(16), 2194-2200.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, *10*(10), 996-998.

Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, btv401.

Edwards, K. J., Wheat, C. G., & Sylvan, J. B. (2011). Under the sea: microbial life in volcanic oceanic crust. *Nature Reviews Microbiology*, **9**(10), 703-712.

Edwards KJ, Becker K & Colwell F (2012). “The deep, dark energy biosphere: intraterrestrial life on earth,” in *Annual Review of Earth and Planetary Sciences*, Vol. 40, ed. Jeanloz R (Palo Alto: Annual Reviews), pp. 551–568.

Ehrlich, H. L., & Newman, D. K. (Eds.). (2008). *Geomicrobiology*. CRC press, New York.

Elbaz-Poulichet, F., & Dupuy, C. (1999). Behaviour of rare earth elements at the freshwater–seawater interface of two acid mine rivers: the Tinto and Odiel (Andalucia, Spain). *Applied Geochemistry*, **14**(8), 1063-1072.

Engelen, B., Ziegelmüller, K., Wolf, L., Köpke, B., Gittel, A., Cypionka, H., Treude, T., Nakagawa, S., Inagaki, F., Lever, M. A., & Steinsbu, B. R. O. (2008). Fluids from the oceanic crust support microbial activities within the deep biosphere. *Geomicrobiology Journal*, **25**(1), 56-66.

Ercole, C., Cacchio, P., Botta, A. L., Centi, V., & Lepidi, A. (2007). Bacterially induced mineralization of calcium carbonate: the role of exopolysaccharides and capsular polysaccharides. *Microscopy and Microanalysis*, **13**(01), 42-50.

Etiope, G., & Sherwood Lollar, B. (2013). Abiotic methane on Earth. *Reviews of Geophysics*, **51**(2), 276-299.

Eynard, A., Del Campillo, M. C., Barrón, V., & Torrent, J. (1992). Use of vivianite (Fe3 (PO4) 2.8 H2O) to prevent iron chlorosis in calcareous soils. *Fertilizer research*, **31**(1), 61-67.

Fanale, F. P., Postawko, S. E., Pollack, J. B., Carr, M. H., & Pepin, R. O. (1992). Mars-Epochal climate change and volatile history. *Mars*, **1**, 1135-1179.

Fang, J., & Zhang, L. (2011). Exploring the deep biosphere. *Science China Earth Sciences*, **54**(2), 157-165.

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, **125**, 1-15.

Felsenstein, J. (1985b). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783-791.

Fernández-Remolar, D. C., Rodriguez, N., Gómez, F., & Amils, R. (2003). Geological record of an acidic environment driven by iron hydrochemistry: The Tinto River system. *Journal of Geophysical Research: Planets* (1991–2012), **108**(E7).

Fernández-Remolar, D. C., Morris, R. V., Gruener, J. E., Amils, R., & Knoll, A. H. (2005). The Rio Tinto Basin, Spain: mineralogy, sedimentary geobiology, and implications for interpretation of outcrop rocks at Meridiani Planum, Mars. *Earth and Planetary Science Letters*, **240**(1), 149-167.

Fernández-Remolar, D. C., Prieto-Ballesteros, O., Rodríguez, N., Gómez, F., Amils, R., Gómez-Elvira, J., & Stoker, C. R. (2008). Underground habitats in the Río Tinto basin: a model for subsurface life habitats on Mars. *Astrobiology*, **8**(5), 1023-1047.

Fernández-Remolar, D. C., Gómez, F., Prieto-Ballesteros, O., Schelble, R. T., Rodríguez, N., & Amils, R. (2008b). Some ecological mechanisms to generate habitability in planetary subsurface areas by chemolithotrophic communities: The Rio Tinto subsurface ecosystem as a model system. Astrobiology, 8(1), 157-173.

Fernández-Remolar, D. C., Preston, L. J., Sanchez-Roman, M., Izawa, M. R., Huang, L., Southam, G., Banerjee, N. R., Osinski, G. R., Flemming, R., Gómez-Ortíz, D., Prieto-Ballesteros, O., Rodríguez, N., Amils, R., & Dyar, M. D. (2012). Carbonate precipitation under bulk acidic conditions as a potential biosignature for searching life on Mars. *Earth and Planetary Science Letters*, *351*, 13-26.

Ferrero, J. J., Alvarez, A. M., Ramírez-Franco, J., Godino, M. C., Bartolomé-Martín, D., Aguado, C., Torres, M., Lujan, R., Ciruela, F., & Sánchez-Prieto, J. (2013). β-Adrenergic Receptors Activate Exchange Protein Directly Activated by cAMP (Epac), Translocate Munc13-1, and Enhance the Rab3A-RIM1α Interaction to Potentiate Glutamate Release at Cerebrocortical Nerve Terminals. *Journal of Biological Chemistry*, **288**(43), 31370-31385.

Findlay, R. H., King, G. M., & Watling, L. (1989). Efficacy of phospholipid analysis in determining microbial biomass in sediments. *Applied and Environmental Microbiology*, **55**(11), 2888-2893.

Finkel, S. E., & Kolter, R. (1999). Evolution of microbial diversity during prolonged starvation. *Proceedings of the National Academy of Sciences*, **96**(7), 4023-4027.

Finkel, S. E. (2006). Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology*, **4**(2), 113-120.

Finster, K. W., Cockell, C. S., Voytek, M. A., Gronstal, A. L., & Kjeldsen, K. U. (2009). Description of Tessaracoccus profundi sp. nov., a deep subsurface actinobacterium isolated from a Chesapeake impact crater drill core (940 m depth). *Antonie Van Leeuwenhoek*, **96**(4), 515-526.

Fisk, M. R., Giovannoni, S. J., & Thorseth, I. H. (1998). Alteration of oceanic volcanic glass: textural evidence of microbial activity. *Science*, **281**(5379), 978-980.

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology*, **20**(4), 406-416.

Folk, R. L. (1993). SEM imaging of bacteria and nannobacteria in carbonate sediments and rocks. *Journal of Sedimentary Research*, 63(5), 990-999.

Formisano, V., Atreya, S., Encrenaz, T., Ignatiev, N., & Giuranna, M. (2004). Detection of methane in the atmosphere of Mars. *Science*, **306**(5702), 1758-1761.

Francis, A. J., Slater, J. M., & Dodge, C. J. (1989). Denitrification in deep subsurface sediments. *Geomicrobiology Journal*, **7**(1-2), 103-116.

França, M. B., Panek, A. D., & Eleutherio, E. C. A. (2007). Oxidative stress and its effects during dehydration. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **146**(4), 621-631.

Fredrickson, J. K., McKinley, J. P., Nierzwicki-Bauer, S. A., White, D. C., Ringelberg, D. B., Rawson, S. A., Li, S., Brockman, F. J., & Bjornstad, B. N. (1995). Microbial community structure and biogeochemistry of Miocene subsurface sediments: implications for long-term microbial survival. *Molecular ecology*, **4**(5), 619-626.

Fredrickson, J. K., & Onstott, T. C. (1996). Microbes deep inside the earth. *Scientific American*, **275**(4), 68-73.

Fredrickson, J. K., & Fletcher, M. (2001). *Biogeochemical and geological significance of subsurface microbiology* (No. PNNL-SA-35352). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

Fredrickson, J. K., & Balkwill, D. L. (2006). Geomicrobial processes and biodiversity in the deep terrestrial subsurface. *Geomicrobiology Journal*, **23**(6), 345-356.

Fredrickson, J. K., Shu-mei, W. L., Gaidamakova, E. K., Matrosova, V. Y., Zhai, M., Sulloway, H. M., Scholten, J. C., Brown, M. G., Balkwill, D. L., & Daly, M. J. (2008). Protein oxidation: key to bacterial desiccation resistance?. *The ISME journal*, **2**(4), 393-403.

Freund, F., Dickinson, J. T., & Cash, M. (2002). Hydrogen in rocks: an energy source for deep microbial communities. *Astrobiology*, **2**(1), 83-92.

Friedmann, E. I. (1980). Endolithic microbial life in hot and cold deserts. *Limits of Life*, **4**, 33-45. Springer Netherlands.

Friedmann, E. I., & Meyer, M. A. (1986). Antarctic cryptoendolithic microbial ecosystem research, 1986-1987. *Antarctic journal of the United States/National Science Foundation*, **22**(5), 240-241.

Früh-Green GL, Connolly JAD, Kelley DS & Grobéty B (2004). Serpentinization of oceanic peridotites: implications for geochemical cycles and biological activity. In: Wilcock W.S.D., editor; DeLong E.F., editor; Kelley D.S., editor; Baross J.A., editor; Cary S.C., editor. *The Subseafloor Biosphere at Mid-Ocean Ridges*. American Geophysical Union; Washington DC: 2004. pp. 119–136. AGU Geophysical Monograph 144.

García-Moyano, A., González-Toril, E., Aguilera, Á., & Amils, R. (2012). Comparative microbial ecology study of the sediments and the water column of the Rio Tinto, an extreme acidic environment. *FEMS microbiology ecology*, **81**(2), 303-314.

Garrido, P., González-Toril, E., García-Moyano, A., Moreno-Paz, M., Amils, R., & Parro, V. (2008). An oligonucleotide prokaryotic acidophile microarray: its validation and its use to monitor seasonal variations in extreme acidic environments with total environmental RNA. *Environmental microbiology*, **10**(4), 836-850.

Gaspar, J. M., & Thomas, W. K. (2013). Assessing the consequences of denoising marker-based metagenomic data. *PLoS One*, **8**(3), e60458.

Gendrin, A., Mangold, N., Bibring, J. P., Langevin, Y., Gondet, B., Poulet, F., Bonello, G., Quantin, C., Mustard, J., Arvidson, R., & LeMouélic, S. (2005). Sulfates in Martian layered terrains: the OMEGA/Mars Express view. *Science*, **307**(5715), 1587-1591.

Gescher, J., Eisenreich, W., Wörth, J., Bacher, A., & Fuchs, G. (2005). Aerobic benzoyl-CoA catabolic pathway in Azoarcus evansii: studies on the non-oxygenolytic ring cleavage enzyme.

Molecular microbiology, **56**(6), 1586-1600.

Gold, T. (1992). The deep, hot biosphere. *Proceedings of the National Academy of Sciences*, **89**(13), 6045-6049.

Gómez, F., & Amils, R. (2002). Life as an environmental transformer. In *The Evolving Sun and its Influence on Planetary Environments* (Vol. 269, p. 339).

Gómez-Ortiz, D., Fernández-Remolar, D. C., Granda, Á., Quesada, C., Granda, T., Prieto-Ballesteros, O., Molina, A., & Amils, R. (2014). Identification of the subsurface sulfide bodies responsible for acidity in Río Tinto source water, Spain. *Earth and Planetary Science Letters*, *391*, 36-41.

González-Toril, E., Gómez, F., Rodrıguez, N., Fernández-Remolar, D., Zuluaga, J., Marın, I., & Amils, R. (2003). Geomicrobiology of the Tinto River, a model of interest for biohydrometallurgy. *Hydrometallurgy*, **71**(1), 301-309.

Gorbushina, A. A., & Krumbein, W. E. (2000). Subaerial microbial mats and their effects on soil and rock. In *Microbial Sediments* (pp. 161-170). Springer Berlin Heidelberg.

Gorbushina, A. A., Beck, A., & Schulte, A. (2005). Microcolonial rock inhabiting fungi and lichen photobionts: evidence for mutualistic interactions. *Mycological Research*, **109**(11), 1288-1296.

Gostinčar, C., Muggia, L., & Grube, M. (2012). Polyextremotolerant black fungi: oligotrophism, adaptive potential, and a link to lichen symbioses. *Frontiers in microbiology*, *3*.

Goyal, K., Walton, L. J., Browne, J. A., Burnell, A. M., & Tunnacliffe, A. (2005). Molecular anhydrobiology: identifying molecules implicated in invertebrate anhydrobiosis. *Integrative and Comparative Biology*, **45**(5), 702-709.

Greaves, M. P., & Wilson, M. J. (1969). The adsorption of nucleic acids by montmorillonite. *Soil Biology and Biochemistry*, **1**(4), 317-323.

Greenwood, N. N., & Earnshaw, A. (2012). *Chemistry of the Elements*. Elsevier.

Grotzinger, J. P., Gupta, S., Malin, M. C., Rubin, D. M., Schieber, J., Siebach, K., Sumner, D. Y., Stack, K. M., Vasavada, A. R., Arvidson, R. E., Calef, F., Edgar, L., Fischer, W. F., Grant, J. A., Griffes, J., Kah, L. C., Lamb, M. P., Lewis, K. W., Mangold, N., Minitti, M. E., Palucis, M., Rice, M., Williams, R. M. E., Yingst, R. A., Blake, D., Blaney, D., Conrad, P., Crisp, J., Dietrich, W. E., Dromart, G., Edgett, K. S., Ewing, R. C., Gellert, R., Hurowitz, J. A., Kocurek, G., Mahaffy, P. R., McBride, M. J., McLennan, S. M., Mischna, M. A., Ming, D., Milliken, R., Newsom, H., Oehler, D., Parker, T. J., Vaniman, D., Wiens, R. C. & Wilson, S. A. (2015). Deposition, exhumation, and paleoclimate of an ancient lake deposit, Gale crater, Mars. Science, **350** (6257). aac7575.

Grube, M., & Berg, G. (2009). Microbial consortia of bacteria and fungi with focus on the lichen symbiosis. *Fungal biology reviews*, **23**(3), 72-85.

Guckert, J. B., Hood, M. A., & White, D. C. (1986). Phospholipid ester-linked fatty acid profile changes during nutrient deprivation of Vibrio cholerae: increases in the trans/cis ratio and

proportions of cyclopropyl fatty acids. *Applied and Environmental Microbiology*, **52**(4), 794-801.

Gueidan, C., Villaseñor, C. R., De Hoog, G. S., Gorbushina, A. A., Untereiner, W. A., & Lutzoni, F. (2008). A rock-inhabiting ancestor for mutualistic and pathogen-rich fungal lineages. *Studies in Mycology*, **61**, 111-119.

Gueidan, C., Ruibal, C., De Hoog, G. S., & Schneider, H. (2011). Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. *Fungal biology*, **115**(10), 987-996.

Guillén, F., Martínez, M. J., Muñoz, C., & Martínez, A. T. (1997). Quinone Redox Cycling in the Ligninolytic FungusPleurotus eryngiiLeading to Extracellular Production of Superoxide Anion Radical. *Archives of biochemistry and biophysics*, **339**(1), 190-199.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., The Human Microbiome Consortium, Petrosino, J. F., Knight, R., & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research*, *21*(3), 494-504.

Haldeman, D. L., Amy, P. S., Ringelberg, D., & White, D. C. (1993). characterisation of the microbiology within a 21 m3 section of rock from the deep subsurface. *Microbial ecology*, **26**(2), 145-159.

Haldeman, D. L., Amy, P. S., White, D. C., & Ringelberg, D. B. (1994). Changes in bacteria recoverable from subsurface volcanic rock samples during storage at 4 ºC. *Applied and environmental microbiology*, **60**(8), 2697-2703.

Hallberg, K. B., & Johnson, D. B. (2001). Biodiversity of acidophilic prokaryotes. *Advances in applied microbiology*, **49**, 37-84.

Hammack, R. N., & Watzlaf, G. R. (1990). The effect of oxygen on pyrite oxidation. *1990 Mining and Reclamation Conference and Exhibition, Charleston WV*.

Hammes, F., Berney, M., & Egli, T. (2010). Cultivation-independent Assessment of Bacterial Viability. *Advances in biochemical engineering/biotechnology*, **124**, 123-150.

Hartman, H. (1975). Speculations on the origin and evolution of metabolism. *Journal of molecular evolution*, **4**(4), 359-370.

Hasan, K., Yildiz, H. B., Sperling, E., Conghaile, P. Ó., Packer, M. A., Leech, D., Hägerhäll, C., & Gorton, L. (2014). Photo-electrochemical communication between cyanobacteria (Leptolyngbia sp.) and osmium redox polymer modified electrodes. *Physical Chemistry Chemical Physics*, **16**(45), 24676-24680.

Heipieper, H. J., Diefenbach, R. U. T. H., & Keweloh, H. (1992). Conversion of cis unsaturated fatty acids to trans, a possible mechanism for the protection of phenol-degrading Pseudomonas putida P8 from substrate toxicity. *Applied and Environmental Microbiology*, **58**(6), 1847-1852.

Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O., & Shanmugam, K. T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied microbiology and biotechnology*, **43**(4), 667-674.

Hellevang, H., Huang, S., & Thorseth, I. H. (2011). The potential for low-temperature abiotic hydrogen generation and a hydrogen-driven deep biosphere. *Astrobiology*, **11**(7), 711-724.

Hernandez, M. E., & Newman, D. K. (2001). Extracellular electron transfer. *Cellular and Molecular Life Sciences CMLS*, **58**(11), 1562-1571.

Hinrichs, K. U., Hayes, J. M., Bach, W., Spivack, A. J., Hmelo, L. R., Holm, N. G., Johnson, C. G., & Sylva, S. P. (2006). Biological formation of ethane and propane in the deep marine subsurface. *Proceedings of the National Academy of Sciences*, **103**(40), 14684-14689.

Hinrichs, K. U., & Inagaki, F. (2012). Downsizing the deep biosphere. *Science*, **338**(6104), 204-205.

Hock, B. (1995, December). Advances in immunochemical detection of microorganisms. In *Annales de biologie clinique* (Vol. 54, No. 6, pp. 243-252).

Hoehler, T. M. (2004). Biological energy requirements as quantitative boundary conditions for life in the subsurface. *Geobiology*, **2**(4), 205-215.

Hoehler, T. M., & Jørgensen, B. B. (2013). Microbial life under extreme energy limitation. *Nature Reviews Microbiology*, **11**(2), 83-94.

Hoefs, J. (2008). *Stable isotope geochemistry*. Springer.

Holben, W. E. (1994). Isolation and purification of bacterial DNA from soil. *Methods of Soil Analysis: Part 2—Microbiological and Biochemical Properties*, (Weaver, R. W., Angle, J. S., Bottomley, P. S., Eds). Soil Science Society of America, Madison, WI.

Holland, H. D. (1973). The oceans; a possible source of iron in iron-formations. Economic *Geology*, **68**(7), 1169-1172.

Horsfield B & Kieft TL (2007). “The GeoBiosphere”. In: *Continental Scientific Drilling: A Decade of Progress and Challenges for the Future*, eds. Harms U, Koeberl H & Zoback MD (Springer-Verlag, Berlin-Heidelberg), pp. 165-182.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**(5), 680-682.

Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of bacteriology*, **180**(18), 4765-4774.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol*, **3**(2), 1-0003.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., Salter, I., Agogué, H., Mary, I., & Galand, P. E. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(15), 6004-6009.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A., Suzuki, M., Takai, K., Delwiche, M., Colwell, F. S., Nealson, K. H., Horikoshi, K., D'Hondt, S., & Jørgensen, B. B. (2006). Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(8), 2815-2820.

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L., & Kukkonen, I. (2011). characterisation of bacterial diversity to a depth of 1500 m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS microbiology ecology*, **77**(2), 295-309.

Jacobs, J. A., & Testa, S. M. (2014). Acid Drainage and Sulfide Oxidation: Introduction. *Acid Mine Drainage, Rock Drainage, and Acid Sulfate Soils: Causes, Assessment, Prediction, Prevention, and Remediation*, 1-8. John Wiley & Sons.

Jakobsen, R., & Postma, D. (1994). In situ rates of sulfate reduction in an aquifer (Rømø, Denmark) and implications for the reactivity of organic matter. *Geology*, *22*(12), 1101-1106.

Jakosky, B. M., & Shock, E. L. (1998). The biological potential of Mars, the early Earth, and Europa. *Journal of Geophysical Research: Planets (1991–2012)*, **103**(E8), 19359-19364.

Johnson, D. B. (2001). Importance of microbial ecology in the development of new mineral technologies. *Hydrometallurgy*, **59**(2), 147-157.

Johnson, H. P., & Pruis, M. J. (2003). Fluxes of fluid and heat from the oceanic crustal reservoir. *Earth and Planetary Science Letters*, **216**(4), 565-574.

Jones, J. G., Davison, W., & Gardener, S. (1984). Iron reduction by bacteria: range of organisms involved and metals reduced. *FEMS microbiology letters*, **21**(1), 133-136.

Jones, A. A., & Bennett, P. C. (2014). Mineral Microniches Control the Diversity of Subsurface Microbial Populations. *Geomicrobiology Journal*, **31**(3), 246-261.

Kageyama, A., Morisaki, K., Ōmura, S., & Takahashi, Y. (2008). Arthrobacter oryzae sp. nov. and Arthrobacter humicola sp. nov. *International journal of systematic and evolutionary microbiology*, **58**(1), 53-56.

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., & D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences*, **109**(40), 16213-16216.

Kämpfer, P., & Kroppenstedt, R. M. (1996). Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Canadian journal of microbiology*, **42**(10), 989-1005.

Kämpfer, P., Lodders, N., Warfolomeow, I., & Busse, H. J. (2009). *Tessaracoccus lubricantis* sp. nov., isolated from a metalworking fluid. *International journal of systematic and evolutionary*

microbiology, **59**(6), 1545-1549.

Kappler, A., Benz, M., Schink, B., & Brune, A. (2004). Electron shuttling via humic acids in microbial iron (III) reduction in a freshwater sediment. *FEMS Microbiology Ecology*, **47**(1), 85-92.

Kashefi, K., & Lovley, D. R. (2003). Extending the upper temperature limit for life. *Science*, **301**(5635), 934-934.

Kaster, A. K., Moll, J., Parey, K., & Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proceedings of the National Academy of Sciences*, **108**(7), 2981-2986.

Kelley, D. S., Baross, J. A., & Delaney, J. R. (2002). Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences*, **30**(1), 385-491.

Kelley DS, Lilley MD & Früh-Green GL (1994). Volatiles in submarine environments: food for life. In: Wilcock W.S.D., editor; DeLong E.F., editor; Kelley D.S., editor; Baross J.A., editor; Cary S.C., editor. *The Subseafloor Biosphere at Mid-Ocean Ridges*. American Geophysical Union; Washington DC: 2004. pp. 167–189. AGU Geophysical Monograph 144.

Kenkel, N. C. (2006). On selecting an appropriate multivariate analysis. *Canadian Journal of Plant Science*, **86**(3), 663-676.

Kerem, Z., Jensen, K. A., & Hammel, K. E. (1999). Biodegradative mechanism of the brown rot basidiomycete Gloeophyllum trabeum: evidence for an extracellular hydroquinone-driven Fenton reaction. *FEBS letters*, **446**(1), 49-54.

Kieft, T. L., Amy, P. S., Brockman, F. J., Fredrickson, J. K., Bjornstad, B. N., & Rosacker, L. L. (1993). Microbial abundance and activities in relation to water potential in the vadose zones of arid and semiarid sites. *Microbial ecology*, **26**(1), 59-78.

Kieft, T. L., Ringelberg, D. B., & White, D. C. (1994). Changes in ester-linked phospholipid fatty acid profiles of subsurface bacteria during starvation and desiccation in a porous medium. *Applied and Environmental Microbiology*, **60**(9), 3292-3299.

Kietäväinen, R., & Purkamo, L. (2015). The origin, source, and cycling of methane in deep crystalline rock biosphere. *Frontiers in microbiology*, **6**.

Kieft, T. L. (2010). Sampling the Deep Sub-Surface Using Drilling and Coring Techniques. In *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 3427-3441). Springer Berlin Heidelberg.

King, G. M. (2006). Nitrate-dependent anaerobic carbon monoxide oxidation by aerobic CO-oxidizing bacteria. *FEMS microbiology ecology*, **56**(1), 1-7.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research*, gks808.

Klingelhöfer, G. R. D. S., Morris, R. V., Bernhardt, B., Schröder, C., Rodionov, D. S., De Souza, P. A., Yen, A., Gellert, R., Evlanov, E. N., Zubkov, B., & Arvidson, R. E. (2004). Jarosite and

hematite at Meridiani Planum from Opportunity's Mössbauer spectrometer. *Science*, **306**(5702), 1740-1745.

Knittel, K., & Boetius, A. (2009). Anaerobic oxidation of methane: progress with an unknown process. *Annual review of microbiology*, **63**, 311-334.

Knoll, A. H., Golubic, S., Green, J., & Swett, K. (1986). Organically preserved microbial endoliths from the Late Proterozoic of East Greenland. *Nature*, **321**, 856-857.

Kobayashi, M., Matsuo, Y., Takimoto, A., Suzuki, S., Maruo, F., & Shoun, H. (1996). Denitrification, a novel type of respiratory metabolism in fungal mitochondrion. *Journal of Biological Chemistry*, **271**(27), 16263-16267.

Kormas, K. A., Smith, D. C., Edgcomb, V., & Teske, A. (2003). Molecular analysis of deep subsurface microbial communities in Nankai Trough sediments (ODP Leg 190, Site 1176). *FEMS Microbiology Ecology*, **45**(2), 115-125.

Korotkova, N., Chistoserdova, L., Kuksa, V., & Lidstrom, M. E. (2002). Glyoxylate regeneration pathway in the methylotroph Methylobacterium extorquens AM1. *Journal of bacteriology*, **184**(6), 1750-1758.

Kostka, J. E., Dalton, D. D., Skelton, H., Dollhopf, S., & Stucki, J. W. (2002). Growth of iron (III)-reducing bacteria on clay minerals as the sole electron acceptor and comparison of growth yields on a variety of oxidized iron forms. *Applied and Environmental Microbiology*, **68**(12), 6256-6262.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, **79**(17), 5112-5120.

Krause, S., Liebetrau, V., Gorb, S., Sánchez-Román, M., McKenzie, J. A., & Treude, T. (2012). Microbial nucleation of Mg-rich dolomite in exopolymeric substances under anoxic modern seawater salinity: New insight into an old enigma. *Geology*, **40**(7), 587-590.

Kreuzer-Martin, H. W., Ehleringer, J. R., & Hegg, E. L. (2005). Oxygen isotopes indicate most intracellular water in log-phase *Escherichia coli* is derived from metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(48), 17337-17341.

Kreuzer-Martin, H. W., Lott, M. J., Ehleringer, J. R., & Hegg, E. L. (2006). Metabolic processes account for the majority of the intracellular water in log-phase *Escherichia coli* cells as revealed by hydrogen isotopes. *Biochemistry*, **45**(45), 13622-13630.

Krsek, M., & Wellington, E. M. H. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *Journal of Microbiological Methods*, **39**(1), 1-16.

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology*, **12**(1), 118-123.

Kuwazaki, S., Takaya, N., Nakamura, A., & Shoun, H. (2003). Formate-forming fungal catabolic pathway to supply electrons to nitrate respiration. *Bioscience, biotechnology, and biochemistry*, **67**(4), 937-939.

Kuykendall, L. D., Roy, M. A., O'neill, J. J., & Devine, T. E. (1988). Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. *International journal of systematic bacteriology*, **38**(4), 358-361.

Laempe, D., Jahn, M., Breese, K., Schägger, H., & Fuchs, G. (2001). Anaerobic metabolism of 3-hydroxybenzoate by the denitrifying bacterium Thauera aromatica. *Journal of bacteriology*, **183**(3), 968-979.

Lamprinou, V., Skaraki, K., Kotoulas, G., Economou-Amilli, A., & Pantazidou, A. (2012). *Toxopsis calypsus gen. nov., sp. Nov.* (Cyanobacteria, Nostocales) from cave 'Francthi', Peloponnese, Greece: a morphological and molecular evaluation. *International journal of systematic and evolutionary microbiology*, **62**(Pt 12), 2870-2877.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*, **82**(20), 6955-6959.

Lane, D. J. (1991). 16S/23S rRNA sequencing. In: *Nucleic acid techniques in bacterial systematics* pp. 115-175. Edited by Stackebrandt, E., and Goodfellow, M. John Wiley and Sons, New York, NY.

Lange, O. L., Kilian, E., & Ziegler, H. (1986). Water vapor uptake and photosynthesis of lichens: performance differences in species with green and blue-green algae as phycobionts. *Oecologia*, **71**(1), 104-110.

Lange, O. L., Meyer, A., Zellner, H., & Heber, U. (1994). Photosynthesis and water relations of lichen soil crusts: field measurements in the coastal fog zone of the Namib Desert. *Functional Ecology*, **8**(2), 253-264.

Lange, O. L., Allan Green, T. G., Melzer, B., Meyer, A., & Zellner, H. (2006). Water relations and CO_2 exchange of the terrestrial lichen *Teloschistes capensis* in the Namib fog desert: Measurements during two seasons in the field and under controlled conditions. *Flora-Morphology, Distribution, Functional Ecology of Plants*, **201**(4), 268-280.

Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*, **79**(14), 4381-4385.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., López, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**(21), 2947-2948.

Leblanc, M., Morales, J. A., Borrego, J., & Elbaz-Poulichet, F. (2000). 4,500-year-old mining pollution in southwestern Spain: long-term implications for modern mining pollution. *Economic Geology*, **95**(3), 655-662.

Lee, A. K., & Newman, D. K. (2003). Microbial iron respiration: impacts on corrosion processes. *Applied microbiology and biotechnology*, **62**(2-3), 134-139.

Lee, D. W., & Lee, S. D. (2008). Tessaracoccus flavescens sp. nov., isolated from marine sediment. *International journal of systematic and evolutionary microbiology*, **58**(4), 785-789.

Leistel, J. M., Marcoux, E., Thieblemont, D., Quesada, C., Sánchez, A., Almodovar, G. R., Pascual, E., & Sáez, R. (1998). The volcanic-hosted massive sulphide deposits of the Iberian Pyrite Belt. Review and preface to the special issue. *Mineralium Deposita* **33**, 2-30.

Lemos, V. P., Costa, M. L. D., Lemos, R. L., & Faria, M. S. G. D. (2007). Vivianite and siderite in lateritic iron crust: an example of bioreduction. *Química Nova*, **30**(1), 36-40.

Lennon, J. T., & Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, **9**(2), 119-130.

Lever, M. A. (2011). Acetogenesis in the energy-starved deep biosphere–a paradox?. *Frontiers in microbiology*, **2**, 1-14.

L'Haridon, S., Reysenbacht, A. L., Glenat, P., Prieur, D., & Jeanthon, C. (1995). Hot subterranean biosphere in a continental oil reservoir. *Nature*, **377**(6546), 223-224.

Liebensteiner, M. G., Tsesmetzis, N., Stams, A. J., & Lomans, B. P. (2014). Microbial redox processes in deep subsurface environments and the potential application of (per) chlorate in oil reservoirs. *Frontiers in microbiology*, **5**, 428.

Lim, Y. W., Baik, K. S., Han, S. K., Kim, S. B., & Bae, K. S. (2003). Burkholderia sordidicola sp. nov., isolated from the white-rot fungus Phanerochaete sordida. *International journal of systematic and evolutionary microbiology*, **53**(5), 1631-1636.

Lin, L. H., Hall, J., Lippmann-Pipke, J., Ward, J. A., Lollar, B. S., DeFlaun, M., Rothmel, R., Moser, D., Gihring, T. M., Mislowack, B., & Onstott, T. C. (2005). Radiolytic H_2 in continental crust: nuclear power for deep subsurface microbial communities. *Geochemistry, Geophysics, Geosystems*, **6**(7), Q07003.

Lin, L. H., Hall, J., Onstott, T. C., Gihring, T., Lollar, B. S., Boice, E., Pratt, L., Lippmann-Pipke, J., & Bellamy, R. E. (2006). Planktonic microbial communities associated with fracture-derived groundwater in a deep gold mine of South Africa. *Geomicrobiology Journal*, **23**(6), 475-497.

Lindahl, T., & Karlstrom, O. (1973). Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry*, **12**(25), 5151-5154.

Lipp, J. S., Morono, Y., Inagaki, F., & Hinrichs, K. U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature*, **454**(7207), 991-994.

Liu, S. V., Zhou, J., Zhang, C., Cole, D. R., Gajdarziska-Josifovska, M., & Phelps, T. J. (1997). Thermophilic Fe (III)-reducing bacteria from the deep subsurface: the evolutionary implications. *Science*, **277**(5329), 1106-1109.

Lollar, B. S., Lacrampe-Couloume, G., Slater, G. F., Ward, J., Moser, D. P., Gihring, T. M., Lin, L. H., & Onstott, T. C. (2006). Unravelling abiogenic and biogenic sources of methane in the Earth's deep subsurface. *Chemical Geology*, **226**(3), 328-339.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*, **30**(5), 434-439.

Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jørgensen, B. B., & Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature*, **484**(7392), 101-104.

López-Archilla, A. I., Marin, I., & Amils, R. (2001). Microbial community composition and ecology of an acidic aquatic environment: the Tinto River, Spain. *Microbial ecology*, **41**(1), 20-35.

Lorite, M. J., Tachil, J., Sanjuán, J., Meyer, O., & Bedmar, E. J. (2000). Carbon Monoxide Dehydrogenase Activity inBradyrhizobium japonicum. *Applied and environmental microbiology*, **66**(5), 1871-1876.

Lovley, D. R., & Phillips, E. J. (1987). Competitive mechanisms for inhibition of sulfate reduction and methane production in the zone of ferric iron reduction in sediments. *Applied and Environmental Microbiology*, *53*(11), 2636-2641.

Lovley, D. R., & Chapelle, F. H. (1995). Deep subsurface microbial processes. *Reviews of Geophysics*, **33**(3), 365-381.

Lovley, D. (2000). Fe (III) and Mn (IV) reduction. In: D.R. Lovley (ed.) *Environmental microbe-metal interactions*, ASM Press, Washington, pp. 3- 30.

Lucey, K. S., & Leadbetter, J. R. (2014). Catechol 2, 3-dioxygenase and other meta-cleavage catabolic pathway genes in the 'anaerobic'termite gut spirochete Treponema primitia. *Molecular ecology*, **23**(6), 1531-1543.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Gangolf, J., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Lüßmann, R., May, M., Nonhoff, B, Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., & Schleifer, K. H. (2004). ARB: a software environment for sequence data. *Nucleic acids research*, **32**(4), 1363-1371.

Ma, W. K., Farrell, R. E., & Siciliano, S. D. (2008). Soil formate regulates the fungal nitrous oxide emission pathway. *Applied and environmental microbiology*, **74**(21), 6690-6696.

Madden, M. E., Bodnar, R. J., & Rimstidt, J. D. (2004). Jarosite as an indicator of water-limited chemical weathering on Mars. *Nature*, **431**(7010), 821-823.

Mallick, N., Rai, L. C., Mohn, F. H., & Soeder, C. J. (1999). Studies on nitric oxide (NO) formation by the green alga *Scenedesmus obliquus* and the diazotrophic cyanobacterium *Anabaena doliolum*. *Chemosphere*, **39**(10), 1601-1610.

Manning, P. G., Prepas, E. E., & Serediak, M. S. (1999). Pyrite and vivianite intervals in the bottom sediments of eutrophic Baptiste Lake, Alberta, Canada. *Canadian mineralogist*, **37**, 593-602.

Markillie, L. M., Varnum, S. M., Hradecky, P., & Wong, K. K. (1999). Targeted mutagenesis by duplication insertion in the radioresistant bacterium Deinococcus radiodurans: radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants. *Journal of bacteriology*, **181**(2), 666-669.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, **24**(3), 133-141.

Marlow, J. J., Steele, J. A., Ziebis, W., Thurber, A. R., Levin, L. A., & Orphan, V. J. (2014). Carbonate-hosted methanotrophy represents an unrecognized methane sink in the deep sea. *Nature communications*, **5**.

Martin, W., & Russell, M. J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **358**(1429), 59-85.

Mason, O. U., Di Meo-Savoie, C. A., Van Nostrand, J. D., Zhou, J., Fisk, M. R., & Giovannoni, S. J. (2008). Prokaryotic diversity, distribution, and insights into their role in biogeochemical cycling in marine basalts. *The ISME journal*, **3**(2), 231-242

Mason, O. U., Nakagawa, T., Rosner, M., Van Nostrand, J. D., Zhou, J., Maruyama, A., Fisk, M. R., & Giovannoni, S. J. (2010). First investigation of the microbiology of the deepest layer of ocean crust. *PLoS One*, **5**(11), e15399.

Maszenan, A. M., Seviour, R. J., Patel, B. K. C., Schumann, P., & Rees, G. N. (1999). *Tessaracoccus bendigoensis* gen. nov., sp. nov., a Gram-positive coccus occurring in regular packages or tetrads, isolated from activated sludge biomass. *International journal of systematic bacteriology*, **49**(2), 459-468.

McCollom, T. M. (1999). Methanogenesis as a potential source of chemical energy for primary biomass production by autotrophic organisms in hydrothermal systems on Europa. *Journal of Geophysical Research: Planets (1991–2012)*, **104**(E12), 30729-30742.

McCollom, T. M., & Amend, J. P. (2005). A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic micro-organisms in oxic and anoxic environments. *Geobiology*, **3**(2), 135-144.

McKay, C.P. (2001) The deep biosphere: lessons for planetary exploration. In Subsurface Microbiology and Biogeochemistry, edited by J.K. Fredickson and M. Fletcher, Wiley-Liss, New York, pp 315–328.

McMahon, P. B., & Chapelle, F. H. (1991). Microbial production of organic acids in aquitard sediments and its role in aquifer geochemistry. *Nature* **349**, 233-235.

McMahon, P. B., Chapelle, F. H., Falls, W. F., & Bradley, P. M. (1992). Role of microbial

processes in linking sandstone diagenesis with organic-rich clays. *Journal of Sedimentary Research*, **62**(1).

McMahon, S., & Parnell, J. (2014). Weighing the deep continental biosphere. *FEMS microbiology ecology*, **87**(1), 113-120.

Mendes, R., Garbeva, P., & Raaijmakers, J. M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS microbiology reviews*, **37**(5), 634-663.

Mesbah, M., Premachandran, U., & Whitman, W. B. (1989). Precise measurement of the G+ C content of deoxyribonucleic acid by high-performance liquid chromatography. *International Journal of Systematic Bacteriology*, **39**(2), 159-167.

Meyer, O., & Schlegel, H. G. (1983). Biology of aerobic carbon monoxide-oxidizing bacteria. *Annual Reviews in Microbiology*, *37*(1), 277-310.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., & Edwards, R. A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, **9**, 386.

Middelburg, J. J. (1989). A simple rate model for organic matter decomposition in marine sediments. *Geochimica et Cosmochimica Acta*, **53**(7), 1577-1581.

Miller, E. J., & Chibnall, A. C. (1932). The proteins of grasses: Preliminary communication. *Biochemical Journal*, 26(2), 392.

Miller, L. T. (1982). Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *Journal of Clinical Microbiology*, **16**(3), 584-586.

Moody, J. B. (1976). Serpentinization: a review. *Lithos*, **9**(2), 125-138.

Moran, M. A. (2010). Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe*, **4**(7), 329-335.

Morgan, L. (2005). Molecular evidence for microbially-mediated sulfur cycling in the deep subsurface of the Witwatersrand Basin, South Africa. *Carleton College, Northfield, Minnesota.*

Morita, R. Y. (1997) Bacteria in Oligotrophic Environments (Chapman & Hall, New York).

Morita, R. Y. (1999). Is H_2 the universal energy source for long-term survival?. *Microbial ecology*, **38**(4), 307-320.

Morono, Y., Terada, T., Nishizawa, M., Ito, M., Hillion, F., Takahata, N., Sano, Y., & Inagaki, F. (2011). Carbon and nitrogen assimilation in deep subseafloor microbial cells. *Proceedings of the National Academy of Sciences*, **108**(45), 18295-18300.

Morowitz HJ (1968) in *Energy Flow in Biology*. New York: Academic Press.

Morse, J. W. (1983). The kinetics of calcium carbonate dissolution and precipitation. *Reviews in Mineralogy and Geochemistry*, **11**(1), 227-264.

Morse, J. W., Arvidson, R. S., & Lüttge, A. (2007). Calcium carbonate formation and dissolution. *Chemical reviews*, **107**(2), 342-381.

Moser, D. P., Onstott, T. C., Fredrickson, J. K., Brockman, F. J., Balkwill, D. L., Drake, G. R., Pfiffner, S. M., White, D. C., Takai, K., Pratt, L. M., Fong, J., Lollar, B. S., Slater, G., Phelps, T. J., Spoelstra, N., Deflaun, M., Southam, G., Welty, A. T., Baker, B. J., & Hoek, J. (2003). Temporal shifts in the geochemistry and microbial community structure of an ultradeep mine borehole following isolation. *Geomicrobiology Journal*, **20**(6), 517-548.

Muggia, L., Gueidan, C., Knudsen, K., Perlmutter, G., & Grube, M. (2013). The lichen connections of black fungi. *Mycopathologia*, *175*(5-6), 523-535.

Muggia, L., Fleischhacker, A., Kopun, T., & Grube, M. (2015). Extremotolerant fungi from alpine rock lichens and their phylogenetic relationships. *Fungal Diversity*, 1-24.

Mukhopadhyay, P. K. (1992). Maturation of organic matter as revealed by microscopic methods: applications and limitations of vitrinite reflectance, and continuous spectral and pulsed laser fluorescence spectroscopy. *Developments in Sedimentology*, **47**, 435-510.

Mumma, M. J., Villanueva, G. L., Novak, R. E., Hewagama, T., Bonev, B. P., DiSanti, M. A., Mandell, A., & Smith, M. D. (2009). Strong release of methane on Mars in northern summer 2003. *Science*, **323**(5917), 1041-1045.

Möhlmann, D. (2011). Three types of liquid water in icy surfaces of celestial bodies. *Planetary and Space Science*, **59**(10), 1082-1086.

Nai, C., Wong, H. Y., Pannenbecker, A., Broughton, W. J., Benoit, I., de Vries, R. P., Gueidan C., & Gorbushina, A. A. (2013). Nutritional physiology of a rock-inhabiting, model microcolonial fungus from an ancestral lineage of the Chaetothyriales (Ascomycetes). *Fungal Genetics and Biology*, **56**, 54-66.

Nai, C. (2014). Rock-inhabiting fungi studied with the aid of the model black fungus Knufia petricola A95 and other related strains (Doctoral dissertation, Freie Universität Berlin, Germany).

Neal, J. L., Allen, G. C., Morse, R. D., & Wolf, D. D. (1983). Nitrate, nitrite, nitrous oxide and oxygen-dependent hydrogen uptake by Rhizobium. *FEMS Microbiology Letters*, **17**(1), 335-338.

Newberry, C. J., Webster, G., Cragg, B. A., Parkes, R. J., Weightman, A. J., & Fry, J. C. (2004). Diversity of prokaryotes and methanogenesis in deep subsurface sediments from the Nankai Trough, Ocean Drilling Program Leg 190. *Environmental Microbiology*, **6**(3), 274-287.

Newman, D. K., & Kolter, R. (2000). A role for excreted quinones in extracellular electron transfer. *Nature*, **405**(6782), 94-97.

Nishio, K., Pornpitra, T., Izawa, S., Nishiwaki-ohkawa, T., Kato, S., Hashimoto, K., & Nakanishi, S. (2015). Electrochemical Detection of Circadian Redox Rhythm in Cyanobacterial

Cells via Extracellular Electron Transfer. *Plant and Cell Physiology*, pcv066.

Oggerin, M., Tornos, F., Rodríguez, N., Del Moral, C., Sánchez-Román, M., & Amils, R. (2013). Specific jarosite biomineralization by *Purpureocillium lilacinum*, an acidophilic fungi isolated from Río Tinto. *Environmental microbiology*, **15**(8), 2228-2237.

Ogram, A (1998). Isolation of nucleic acids from environmental samples. In Burlage, R. S. (Ed.). (1998). *Techniques in microbial ecology*. Oxford University Press.

Okland, I., Huang, S., Thorseth, I. H., & Pedersen, R. B. (2014). Formation of H_2, CH_4 and N-species during low-temperature experimental alteration of ultramafic rocks. *Chemical Geology*, **387**, 22-34.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. *Community ecology package*.

Olías, M., & Nieto, J. M. (2012). El impacto de la minería en los ríos Tinto y Odiel a lo largo de la historia. *Revista de la Sociedad Geológica de España*, **25**(3), 177-192.

Olías, M., & Nieto, J. M. (2014). Comment on "Identification of the subsurface sulfide bodies responsible for acidity in Río Tinto source water, Spain" by Gómez-Ortiz et al.(Earth Planet. Sci. Lett. 391 (2014) 36–41). *Earth and Planetary Science Letters*, **403**(Complete), 456-458.

Omar, G. I., Onstott, T. C., & Hoek, J. (2003). The origin of deep subsurface microbial communities in the Witwatersrand Basin, South Africa as deduced from apatite fission track analyses. *Geofluids*, **3**(1), 69-80.

Onstott, T. C., Phelps, T. J., Colwell, F. S., Ringelberg, D., White, D. C., Boone, D. R., McKinley, J. P., Stevens, T. O., Long, P. E., Balkwill, D. L., Griffin, W. T., & Kieft, T. (1998). Observations pertaining to the origin and ecology of microorganisms recovered from the deep subsurface of Taylorsville Basin, Virginia. *Geomicrobiology Journal*, **15**(4), 353-385.

Onstott, T. C., Moser, D. P., Pfiffner, S. M., Fredrickson, J. K., Brockman, F. J., Phelps, T. J., White, D. C., Peacock, A., Balkwill, D., Hoover, R., Krumholz, L. R., Borscik, M., Kieft, T. K., & Wilson, R. (2003). Indigenous and contaminant microbes in ultradeep mines. *Environmental Microbiology*, **5**(11), 1168-1191.

Onstott, T. C., Magnabosco, C., Aubrey, A. D., Burton, A. S., Dworkin, J. P., Elsila, J. E., Grunsfeld, S., Cao, B. H., Hein, J. E, Glavin, D. P., Kieft, T. L., Silver, B. J., Phelps, T. J., van Heerden, E., Opperman, D. J., & Bada, J. L. (2014). Does aspartic acid racemization constrain the depth limit of the subsurface biosphere?. *Geobiology*, **12**(1), 1-19.

Oparin, A. I. (1924) Proiskhozhdenie zhizny. Moscow. Izd. Mosk. Rabochii. English translation by Synge A (1967). In: Bernal JD (ed) The origin of life. Weidenfeld & Nicolson, London, pp 199–234.

Oren, A., & Shilo, M. (1979). Anaerobic heterotrophic dark metabolism in the cyanobacterium Oscillatoria limnetica: sulfur respiration and lactate fermentation. *Archives of Microbiology*, **122**(1), 77-84.

Orcutt, B. N., Sylvan, J. B., Knab, N. J., & Edwards, K. J. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiology and Molecular Biology Reviews*, **75**(2), 361-422.

Orphan, V. J., Hinrichs, K. U., Ussler, W. I. I. I., Paull, C. K., Taylor, L. T., Sylva, S. P., Hayes, J. M., & DeLong, E. F. (2001). Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Applied and Environmental Microbiology*, *67*(4), 1922-1934.

Orsi, W. D., Edgcomb, V. P., Christman, G. D., & Biddle, J. F. (2013). Gene expression in the deep biosphere. *Nature*, **499**(7457), 205-208.

Osburn, M. R., LaRowe, D. E., Momper, L. M., & Amend, J. P. (2014). Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Frontiers in microbiology*, **5**, 610.

Ottow, J. C. G., & Glathe, H. (1971). Isolation and identification of iron-reducing bacteria from gley soils. *Soil Biology and Biochemistry*, **3**(1), 43-55.

Oze, C., & Sharma, M. (2007). Serpentinization and the inorganic synthesis of H_2 in planetary surfaces. *Icarus*, **186**(2), 557-561.

Parkes, R. J., Cragg, B. A., Bale, S. J., Getlifff, J. M., Goodman, K., Rochelle, P. A., Fry, J. C., Weightman, A. J., & Harvey, S. M. (1994). Deep bacterial biosphere in Pacific Ocean sediments. *Nature*, **371**(6496), 410-413.

Parkes, R. J., Cragg, B. A., & Wellsbury, P. (2000). Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeology Journal*, **8**(1), 11-28.

Parkes, R. J., Webster, G., Cragg, B. A., Weightman, A. J., Newberry, C. J., Ferdelman, T. G., Kallmeyer, J., Jørgensen, B. B., Aiello, I. W., & Fry, J. C. (2005). Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature*, **436**(7049), 390-394.

Parkes, R. J., Cragg, B. A., Banning, N., Brock, F., Webster, G., Fry, J. C., Hornibrook, E., Pancost, R. D., Kelly, S., Knab, N., Jørgensen, B. B., Rinna, J., & Weightman, A. J. (2007). Biogeochemistry and biodiversity of methane cycling in subsurface marine sediments (Skagerrak, Denmark). *Environmental Microbiology*, **9**(5), 1146-1161.

Parro, V., Rodríguez-Manfredi, J. A., Briones, C., Compostizo, C., Herrero, P. L., Vez, E., Sebastian, E., Moreno-Paz, M., García-Villadangos, M., Fernández-Calvo, P., González-Toril, E., Pérez-Mercader, J., Fernández-Remolar, D., & Gómez-Elvira, J. (2005). Instrument development to search for biomarkers on Mars: terrestrial acidophile, iron-powered chemolithoautotrophic communities as model systems. *Planetary and Space Science*, **53**(7), 729-737.

Parro, V., Fernández-Calvo, P., Rodríguez Manfredi, J. A., Moreno-Paz, M., Rivas, L. A., García-Villadangos, M., Bonaccorsi, R., González-Pastor, J. E., Prieto-Ballesteros, O., Schuerger, A. C., Davidson, M., Gómez-Elvira, J., & Stoker, C. R. (2008). SOLID2: an antibody array-based life-detector instrument in a Mars drilling simulation experiment (MARTE). *Astrobiology*, **8**(5), 987-999.

Parro, V. (2010). Antibody microarrays for environmental monitoring. In *Handbook of hydrocarbon and lipid microbiology* (pp. 2699-2710). Springer Berlin Heidelberg.

Parro, V., de Diego-Castilla, G., Moreno-Paz, M., Blanco, Y., Cruz-Gil, P., Rodríguez-Manfredi, J. A., Fernández-Remolar, D., Gómez, F., Gómez, M. J., Rivas, L. A., Demergasso, C., Echeverría, A., Urtuvia, V. N., Ruíz-Bermejo, M., García-Villadangos, M., Postigo, M., Sánchez-Román, M., Chong-Díaz, G., & Gómez-Elvira, J. (2011). A microbial oasis in the hypersaline Atacama subsurface discovered by a life detector chip: implications for the search for life on Mars. *Astrobiology*, **11**(10), 969-996.

Pedersen, K. (2000). Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS microbiology letters*, **185**(1), 9-16.

Pedersen, K. (2013). Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H2. *The ISME journal*, **7**(4), 839-849.

Pedersen K (2014). "Microbial life in terrestrial hard rock environments". In *Microbial Life of the Deep Biosphere*, eds: Kallmeyer J & Wagner D (Walter de Gruyter GmbH, Berlin-Boston), pp. 63-82.

Peterson, C. A., & Cowling, E. B. (1973). Influence of various initial moisture contents on decay of sitka spruce and sweetgum sapwood by *Polyporus versicolor* in the soil-block test. *Phytopathology*, **63**, 235-237.

Pentecost, A. (2005). Organisms associated with travertine. In *Travertine* (p. 161). Springer Science & Business Media.

Percheron, G., Bernet, N., & Moletta, R. (1999). Interactions between methanogenic and nitrate reducing bacteria during the anaerobic digestion of an industrial sulfate rich wastewater. *FEMS microbiology ecology*, **29**(4), 341-350.

Pernthaler, A., Pernthaler, J., & Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**(6), 3094-3101.

Philipp, B., Kemmler, D., Hellstern, J., Gorny, N., Caballero, A., & Schink, B. (2002). Anaerobic degradation of protocatechuate (3, 4-dihydroxybenzoate) by Thauera aromatica strain AR-1. *FEMS microbiology letters*, **212**(1), 139-143.

Phoenix, V. R., & Konhauser, K. O. (2008). Benefits of bacterial biomineralization. *Geobiology*, **6**(3), 303-308.

Pinedo-Vara. I. (1963). Piritas de Huelva. Su historia, minería y aprovechamiento . *Summa. Madrid*, 1003 p.

Pini, F., Galardini, M., Bazzicalupo, M., & Mengoni, A. (2011). Plant-bacteria association and symbiosis: are there common genomic traits in Alphaproteobacteria?. *Genes*, **2**(4), 1017-1032.

Pintado, A., & Sancho, L. G. (2002). Ecological significance of net photosynthesis activation by water vapour uptake in Ramalina capitata from rain-protected habitats in central Spain. *The*

Lichenologist, **34**(05), 403-413.

Pisciotta, J. M., Zou, Y., & Baskakov, I. V. (2010). Light-dependent electrogenic activity of cyanobacteria. *PloS one*, **5**(5), e10821.

Pisciotta, J. M., Zou, Y., & Baskakov, I. V. (2011). Role of the photosynthetic electron transfer chain in electrogenic activity of cyanobacteria. *Applied microbiology and biotechnology*, **91**(2), 377-385.

Pitt, J. I., & Christian, J. H. B. (1968). Water relations of xerophilic fungi isolated from prunes. *Applied Microbiology*, **16**(12), 1853-1858.

Poelchau, H. S., Baker, D. R., Hantschel, T., Horsfield, B., & Wygrala, B. (1997). Basin simulation and the design of the conceptual basin model. In *Petroleum and basin evolution* (pp. 3-70). Springer Berlin Heidelberg.

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and environmental Microbiology*, **64**(10), 3724-3730.

Porras-Martín, J., Nieto-Lopez-Guerrero, P., Álvarez-Fernández, C., Fernández-Uría, A., & Gimeno, M. V. (1985). Sistemas acuíferos de la cuenca del Guadalquivir. In *Calidad y contaminación de las aguas subterráneas en España. Informe de síntesis. Tomo II* (p 161). *Anejos.* Instituto Geológico y Minero de España.

Porter, K. G., & Feig, Y. S. (1980). The use of DAPI for identifying and counting aquatic microflora1. *Limnology and oceanography*, **25**(5), 943-948.

Postma, D. (1981). Formation of siderite and vivianite and the pore-water composition of a recent bog sediment in Denmark. *Chemical Geology*, **31**, 225-244.

Potts, M. (1994). Desiccation tolerance of prokaryotes. *Microbiological reviews*, **58**(4), 755.

Potts, M. (1999). Mechanisms of desiccation tolerance in cyanobacteria. *European Journal of Phycology*, **34**(04), 319-328.

Potts, M., Slaughter, S. M., Hunneke, F. U., Garst, J. F., & Helm, R. F. (2005). Desiccation tolerance of prokaryotes: application of principles to human cells. *Integrative and comparative biology*, **45**(5), 800-809.

Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A., & Alm, E. J. (2013). Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Applied and environmental microbiology*, **79**(21), 6593-6603.

Price, P. B., & Sowers, T. (2004). Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(13), 4631-4636.

Priscu, J. C., Adams, E. E., Lyons, W. B., Voytek, M. A., Mogk, D. W., Brown, R. L., McKay, C. P., Takacs, C., Welch, K. A., Wolf, C. F., Kirshtein, J. D., & Avci, R. (1999). Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, **286**(5447), 2141-2144.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, **35**(21), 7188-7196.

Pruesse, E., Peplies, J., & Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, **28**(14), 1823-1829.

Prussin, A. J., Marr, L. C., & Bibby, K. J. (2014). Challenges of studying viral aerosol metagenomics and communities in comparison with bacterial and fungal aerosols. *FEMS microbiology letters*, **357**(1), 1-9.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., & Itävaara, M. (2014). Heterotrophic Communities Supplied by Ancient Organic Carbon Predominate in Deep Fennoscandian Bedrock Fluids. *Microbial ecology*, 1-14.

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., & Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biol*, **14**(5), R51.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, gks1219.

Quesada, C. (1996). Estructura del sector español de la Faja Pirítica: implicaciones para la exploración de yacimientos. *Boletín Geológico Minero* **107**(3-4), 65-78.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F., & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, **6***(*9), 639-641.

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, **12**(1), 38.

Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P. A., Andersen, G. L., & Sani, R. K. (2010). Microbial and mineralogical characterisations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microbial ecology*, **60**(3), 539-550.

Rebecchi, L., Altiero, T., & Guidetti, R. (2007). Anhydrobiosis: the extreme limit of desiccation tolerance. *Invertebrate Survival Journal*, **4**, 65-81.

Reeder, J., & Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature methods*, **7**(9), 668-669.

Rivas, L. A., García-Villadangos, M., Moreno-Paz, M., Cruz-Gil, P., Gómez-Elvira, J., & Parro, V. (2008). A 200-antibody microarray biochip for environmental monitoring: searching for universal microbial biomarkers through immunoprofiling. *Analytical chemistry*, **80**(21), 7970-7979.

Rivas, L. A., Aguirre, J., Blanco, Y., González-Toril, E., & Parro, V. (2011). Graph-based deconvolution analysis of multiplex sandwich microarray immunoassays: applications for

environmental monitoring. *Environmental microbiology*, **13**(6), 1421-1432.

Roels JA (1983) in: *Energetics and kinetics in Biotechnology.* Amsterdam, Elsevier, Ch 1-5.

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., Daroub, S. H., Camargo, F. A., Farmerie, W. G., & Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, **1**(4), 283-290.

Röling, W. F., Head, I. M., & Larter, S. R. (2003). The microbiology of hydrocarbon degradation in subsurface petroleum reservoirs: perspectives and prospects. *Research in Microbiology*, **154**(5), 321-328.

Rossi, F., & De Philippis, R. (2015). Role of Cyanobacterial Exopolysaccharides in Phototrophic Biofilms and in Complex Microbial Mats. *Life*, **5**(2), 1218-1238.

Ruibal, C., Gueidan, C., Selbmann, L., Gorbushina, A. A., Crous, P. W., Groenewald, J. Z., Muggia, L., Grube, M., Isola, D., Schoch, C. L., Stanley, J. T., Lutzoni, F., & De Hoog, G. S. (2009). Phylogeny of rock-inhabiting fungi related to Dothideomycetes. *Studies in Mycology*, *64*, 123-133.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**(4), 406-425.

Sahl, J. W., Schmidt, R., Swanner, E. D., Mandernack, K. W., Templeton, A. S., Kieft, T. L., Smith, R. L., Sanford, W. E., Callaghan, R. L., Mitton, J. B., & Spear, J. R. (2008). Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Applied and environmental microbiology*, **74**(1), 143-152.

Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., Harkins, T. T., Cookson, B. T., & Hoffman, N. G. (2014). Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. *Applied and environmental microbiology*, **80**(24), 7583-7591.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology*, **12**(1), 87.

Sánchez-Alcalá, I., del Campillo, M. C., Barrón, V., & Torrent, J. (2012). Pot evaluation of synthetic nanosiderite for the prevention of iron chlorosis. *Journal of the Science of Food and Agriculture*, **92**(9), 1964-1973.

Sánchez-Román, M., Vasconcelos, C., Schmid, T., Dittrich, M., McKenzie, J. A., Zenobi, R., & Rivadeneyra, M. A. (2008). Aerobic microbial dolomite at the nanometer scale: Implications for the geologic record. *Geology*, **36**(11), 879-882.

Sánchez-Román, M., Romanek, C. S., Fernández-Remolar, D. C., Sánchez-Navas, A., McKenzie, J. A., Pibernat, R. A., & Vasconcelos, C. (2011). Aerobic biomineralization of Mg-rich carbonates: Implications for natural environments. *Chemical Geology*, **281**(3), 143-150.

Sánchez-Román, M., Fernández-Remolar, D., Amils, R., Sánchez-Navas, A., Schmid, T., San

Martin-Uriz, P., Rodríguez, N., McKenzie, J. A., & Vasconcelos, C. (2014). Microbial mediated formation of Fe-carbonate minerals under extreme acidic conditions. *Scientific reports*, **4**.

Santelli, C. M., Orcutt, B. N., Banning, E., Bach, W., Moyer, C. L., Sogin, M. L., Staudigel, H., & Edwards, K. J. (2008). Abundance and diversity of microbial life in ocean crust. *Nature*, **453**(7195), 653-656.

Santelli, C. M., Edgcomb, V. P., Bach, W., & Edwards, K. J. (2009). The diversity and abundance of bacteria inhabiting seafloor lavas positively correlate with rock alteration. *Environmental microbiology*, **11**(1), 86-98.

Sanz, J. L., Rodríguez, N., Díaz, E. E., & Amils, R. (2011). Methanogenesis in the sediments of Rio Tinto, an extreme acidic river. *Environmental microbiology*, **13**(8), 2336-2341.

Schippers, A., Neretin, L. N., Kallmeyer, J., Ferdelman, T. G., Cragg, B. A., Parkes, R. J., & Jørgensen, B. B. (2005). Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature*, **433**(7028), 861-864.

Schippers, A., Kock, D., Höft, C., Köweker, G., & Siegert, M. (2012). Quantification of microbial communities in subsurface marine sediments of the Black Sea and off Namibia. *Frontiers in microbiology*, **3**, 16.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic acids research*, gku1341.

Schleifer K. H. & Seidl P. H. (1985). Chemical composition and structure of murein. In: *Chemical methods in bacterial systematics* pp 201–219. Edited by Goodfellow M, Minnikin D.E. Academic Press, London.

Schlegel, K., Leone, V., Faraldo-Gómez, J. D., & Müller, V. (2012). Promiscuous archaeal ATP synthase concurrently coupled to Na^+ and H^+ translocation. *Proceedings of the National Academy of Sciences*, **109**(3), 947-952.

Schloss, P., Westcott, S.L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesnievski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, **75**(23), 7537-7541.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS one*, **6**(12), e27310.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**(6), 863-864.

Schönhuber, W., Fuchs, B., Juretschko, S., & Amann, R. (1997). Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Applied and Environmental Microbiology*, **63**(8), 3268-3273.

Schrenk, M. O., Huber, J. A., & Edwards, K. J. (2010). Microbial provinces in the subseafloor. *Annual Review of Marine Science*, **2**, 279-304.

Schuchmann, K., & Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, **12**, 809-821.

Schulte, M., Blake, D., Hoehler, T., & McCollom, T. (2006). Serpentinization and its implications for life on the early Earth and Mars. *Astrobiology*, **6**(2), 364-376.

Schultze-Lam, S., Fortin, D., Davis, B. S., & Beveridge, T. J. (1996). Mineralization of bacterial surfaces. *Chemical Geology*, **132**(1), 171-181.

Schumann, P. (2011). Peptidoglycan Structure. *Methods in Microbiology*, **38**, 101-129.

Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., & Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Molecular systems biology*, **9**(1).

Sharma, A., Scott, J. H., Cody, G. D., Fogel, M. L., Hazen, R. M., Hemley, R. J., & Huntress, W. T. (2002). Microbial activity at gigapascal pressures. *Science*, **295**(5559), 1514-1516.

Shoun, H., Kim, D. H., Uchiyama, H., & Sugiyama, J. (1992). Denitrification by fungi. *FEMS Microbiology Letters*, **94**(3), 277-281.

Silva, J.B., Oliveira, J.T., & Ribeiro, A. (1990). Structural outline of the South Portuguese Zone. In Dallmeyer, R.D., Martinez García, E. (Eds.), PreMesozoic Geology of Iberia: Heidelberg, Springer Verlag, pp. 348-362.

Sim, K., Cox, M. J., Wopereis, H., Martin, R., Knol, J., Li, M. S., ... & Kroll, J. S. (2012). Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS One*, **7**(3), e32543.

Slade, D., & Radman, M. (2011). Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiology and Molecular Biology Reviews*, **75**(1), 133-191.

Slater, G. F., Lippmann-Pipke, J., Moser, D. P., Reddy, C. M., Onstott, T. C., Lacrampe-Couloume, G. & Lollar, B. S.(2006) ^{14}C in methane and DIC in the deep terrestrial subsurface: Implications for microbial methanogenesis. *Geomicrobiology Journal* **23:453**–462.

Smith, J. L. (2004). The physiological role of ferritin-like compounds in bacteria. *Critical reviews in microbiology*, **30**(3), 173-185.

Squyres, S. W., & Kasting, J. F. (1994). Early Mars: How warm and how wet?. *Science*, **265**(5173), 744-749.

Squyres, S. W., Grotzinger, J. P., Arvidson, R. E., Bell, J. F., Calvin, W., Christensen, P. R., Clark, B. C., Crisp, J. A., Farrand, W. H., Herkenhoff, K. E., Johnson, J. R, Klingelhöfer, G., Knoll, A. H., McLennan, S. M., McSween, H. Y., Morris, R. V., Rice, J. W., Rieder, R. & Soderblom, L. A. (2004). In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science*, **306**(5702), 1709-1714.

Squyres, S. W., Knoll, A. H., Arvidson, R. E., Clark, B. C., Grotzinger, J. P., Jolliff, B. L., McLennan, S. M., Tosca, M., Bell III, J. M., Calvin W. M., Farrand W. H., Glotch T. D., Golombek, M. P., Herkenhoff, K. E., Johnson, J. R., Klingelhöfer, G., McSween, H. Y., & Yen, A. S. (2006). Two years at Meridiani Planum: results from the Opportunity Rover. Science, 313(5792), 1403-1407.

Stackebrandt, E., Schumann, P., Schaal, K. P., & Weiss, N. (2002). *Propionimicrobium* gen. nov., a new genus to accommodate *Propionibacterium lymphophilum* (Torrey 1916) Johnson and Cummins 1972, 1057AL as *Propionimicrobium lymphophilum* comb. nov. *International journal of systematic and evolutionary microbiology*, **52**(6), 1925-1927.

Stahl, D. A. (1991). Development and application of nucleic acid probes. *Nucleic acid techniques in bacterial systematics*.

Stal, L. J., & Moezelaar, R. (1997). Fermentation in cyanobacteria. *FEMS microbiology reviews*, **21**(2), 179-211.

Stal, L. J. (2007). Cyanobacterial mats and stromatolites. In Whitton, B. A., & Potts, M. (Eds.). *Ecology of Cyanobacteria* (pp. 61-120). Kluwer Academic Publishers.

Stal, L. J. (2012). Cyanobacterial mats and stromatolites. In Whitton, B. A., & Potts, M. (Eds.). *Ecology of Cyanobacteria II* (pp. 65-125). Springer Netherlands.

Staudigel, H., Furnes, H., McLoughlin, N., Banerjee, N. R., Connell, L. B., & Templeton, A. (2008). 3.5 billion years of glass bioalteration: Volcanic rocks as a basis for microbial life?. *Earth-Science Reviews*, **89**(3), 156-176.

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). characterisation of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of bacteriology*, **178**(3), 591-599.

Sterflinger, K., & Krumbein, W. E. (1997). Dematiaceous fungi as a major agent for biopitting on Mediterranean marbles and limestones. *Geomicrobiology Journal*, **14**(3), 219-230.

Sterflinger, K. (2000). Fungi as geologic agents. *Geomicrobiology Journal*, **17**(2), 97-124.

Stevens, T. O., & McKinley, J. P. (1995). Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science*, **270**(5235), 450-455.

Stewart, F. J., Ottesen, E. A., & DeLong, E. F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME journal*, **4**(7), 896-907.

Stewart, F. J., Ulloa, O., & DeLong, E. F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental microbiology*, **14**(1), 23-40.

Stoker, C. R., Cannon, H. N., Dunagan, S. E., Lemke, L. G., Glass, B. J., Miller, D., Gómez-Elvira, J. Davis, K., Zavaleta, J., Winterholler, A., Roman, M., Rodríguez-Manfredi, J. A., Bonaccorsi, R., Bell, M. S., Brown, A., Battler, M., Chen, B., Cooper, G., Davidson, M., Fernández-

Remolar, D., González-Pastor, E., Heldmann, J. L., Martínez-Frías, J., Parro, V., Prieto-Ballesteros, O., Sutter, B., Schuerger, A. C., Shutt, J. & Rull, F. (2008). The 2005 MARTE robotic drilling experiment in Río Tinto, Spain: objectives, approach, and results of a simulated mission to search for life in the martian subsurface. *Astrobiology*, **8**(5), 921-945.

Stouthamer, A. H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, **39**(1), 545-565.

Stouthamer AH (1979) in *International Review of Biochemistry and Microbial Biochemistry* Vol. 21 (ed. Quayle JR) 1-47. Univ. Park Press.

Straub, K. L., Benz, M., Schink, B., & Widdel, F. (1996). Anaerobic, nitrate-dependent microbial oxidation of ferrous iron. *Applied and Environmental Microbiology*, **62**(4), 1458-1460.

Sturms, R., DiSpirito, A. A., & Hargrove, M. S. (2011). Plant and cyanobacterial hemoglobins reduce nitrite to nitric oxide under anoxic conditions. *Biochemistry*, **50**(19), 3873-3878.

Stuart, R. K., Mayali, X., Lee, J. Z., Everroad, R. C., Hwang, M., Bebout, B. M., Weber, P. K., Pett-Ridge, J., & Thelen, M. P. (2015). Cyanobacterial reuse of extracellular organic carbon in microbial mats. *The ISME journal*.

Takasaki, K., Shoun, H., Yamaguchi, M., Takeo, K., Nakamura, A., Hoshino, T., & Takaya, N. (2004). Fungal ammonia fermentation, a novel metabolic mechanism That couples the dissimilatory and assimilatory pathways of both nitrate and ethanol. Role of acetyl-CoA synthetase in anaerobic ATP synthesis. *Journal of Biological Chemistry*, **279**(13), 12414-12420.

Talbot, H. M., Rohmer, M., & Farrimond, P. (2007). Rapid structural elucidation of composite bacterial hopanoids by atmospheric pressure chemical ionisation liquid chromatography/ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry*, **21**(6), 880-892.

Tamaoka, J., & Komagata, K. (1984). Determination of DNA base composition by reversed-phase high-performance liquid chromatography. *FEMS microbiology letters*, 25(1), 125-128.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, **30**(12), 2725-2729.

Teske, A., & Sørensen, K. B. (2007). Uncultured archaea in deep marine subsurface sediments: have we caught them all?. *The ISME Journal*, **2**(1), 3-18.

Teske, A., Callaghan, A. V., & LaRowe, D. E. (2014). Biosphere frontiers of subsurface life in the sedimented hydrothermal system of Guaymas Basin. *Frontiers in microbiology*, **5**, 362.

Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M., & Michel, G. (2011). Environmental and gut bacteroidetes: the food connection. *Frontiers in microbiology*, **2**.

Tijhuis, L., Van Loosdrecht, M. C., & Heijnen, J. J. (1993). A thermodynamically based correlation for maintenance Gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnology and bioengineering*, **42**(4), 509-519.

Tikhonov, M., Leach, R. W., & Wingreen, N. S. (2015). Interpreting 16S metagenomic data

without clustering to achieve sub-OTU resolution. *The ISME journal*, **9**(1), 68-80.

Tindall, B. J. (1990). A Comparative Study of the Lipid Composition of *Halobacterium saccharovorum* from Various Sources. *Systematic and Applied Microbiology*, **13**(2), 128-130.

Tindall, B. J. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS microbiology letters*, **66**(1), 199-202.

Tindall, B. J., Sikorski, J., Smibert, R. M., & Krieg, N. R. (2007). Phenotypic characterisation and the principles of comparative systematics. In Methods for General and Molecular Microbiology, 3rd edn, pp. 330–393. Edited by C. A. Reddy, T. J. Beveridge, J. A. Breznak, G. Marzluf, T. M. Schmidt & L. R. Snyder. Washington, DC: American Society for Microbiology.

Tornos, F. (2006). Environment of formation and styles of volcanogenic massive sulfides: the Iberian Pyrite Belt. *Ore Geology Reviews*, **28**(3), 259-307.

Tornos, F. (2008). La geología y metalogenía de la faja pirítica ibérica. *Macla*, **10**, 13-23.

Tornos, F., Velasco, F., Menor-Salván, C., Delgado, A., Slack, J. F., & Escobar, J. M. (2014). Formation of recent Pb-Ag-Au mineralization by potential sub-surface microbial activity. *Nature communications*, **5**, 4600.

Torsvik, V. L. (1980). Isolation of bacterial DNA from soil. *Soil Biology and Biochemistry*, **12**(1), 15-21.

Trevors, J. T. (2002). The subsurface origin of microbial life on the Earth. *Research in microbiology*, **153**(8), 487-491.

Troshina, O., Serebryakova, L., Sheremetieva, M., & Lindblad, P. (2002). Production of H 2 by the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743 during fermentation. *International Journal of Hydrogen Energy*, **27**(11), 1283-1289.

Trubitsin, B. V., Ptushenko, V. V., Koksharova, O. A., Mamedov, M. D., Vitukhnovskaya, L. A., Grigor'ev, I. A., Yu, A., Semenov, A. Y., & Tikhonov, A. N. (2005). EPR study of electron transport in the cyanobacterium Synechocystis sp. PCC 6803: oxygen-dependent interrelations between photosynthetic and respiratory electron transport chains. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, **1708**(2), 238-249.

Tuross, N., & Stathoplos, L. (1993). Ancient proteins in fossil bones. *Methods in Enzymology*, **224**, 121-129.

Turner, T. R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A., & Poole, P. S. (2013). Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *The ISME journal*, **7**(12), 2248-2258.

Ulrich, G. A., Martino, D., Burger, K., Routh, J., Grossman, E. L., Ammerman, J. W., & Suflita, J. M. (1998). Sulfur cycling in the terrestrial subsurface: commensal interactions, spatial scales, and microbial heterogeneity. *Microbial ecology*, **36**(2), 141-151.

Utgikar, V. P., Harmon, S. M., Chaudhary, N., Tabak, H. H., Govind, R., & Haines, J. R.

(2002). Inhibition of sulfate-reducing bacteria by metal sulfide formation in bioremediation of acid mine drainage. *Environmental toxicology*, **17**(1), 40-48.

van Baarlen, P., Kleerebezem, M., & Wells, J. M. (2013). Omics approaches to study host–microbiota interactions. *Current opinion in microbiology*, **16**(3), 270-277.

van der Lelie, D., Taghavi, S., McCorkle, S. M., Li, L. L., Malfatti, S. A., Monteleone, D., Donohoe, B. S., Ding, S., Adney, W. S., Himmel, M. E., & Tringe, S. G. (2012). The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PloS one*, *7*(5), e36740.

van Waasbergen, L. G., Balkwill, D. L., Crocker, F. H., Bjornstad, B. N., & Miller, R. V. (2000). Genetic Diversity among Arthrobacter Species Collected across a Heterogeneous Series of Terrestrial deep subsurface Sediments as Determined on the Basis of 16S rRNA and recA Gene Sequences. *Applied and environmental microbiology*, **66**(8), 3454-3463.

Valli, K., & Gold, M. H. (1991). Degradation of 2, 4-dichlorophenol by the lignin-degrading fungus Phanerochaete chrysosporium. *Journal of bacteriology*, **173**(1), 345-352.

Valli, K., Brock, B. J., Joshi, D. K., & Gold, M. H. (1992). Degradation of 2, 4-dinitrotoluene by the lignin-degrading fungus Phanerochaete chrysosporium. *Applied and environmental microbiology*, **58**(1), 221-228.

Van Geen, A., Adkins, J. F., Boyle, E. A., Nelson, C. H., & Palanques, A. (1997). A 120-yr record of widespread contamination from mining of the Iberian pyrite belt. *Geology*, **25**(4), 291-294.

Van Bodegom, P. M., Scholten, J. C., & Stams, A. J. (2004). Direct inhibition of methanogenesis by ferric iron. *FEMS Microbiology Ecology*, **49**(2), 261-268.

Van Lith, Y., Warthmann, R., Vasconcelos, C., & Mckenzie, J. A. (2003). Sulphate-reducing bacteria induce low-temperature Ca-dolomite and high Mg-calcite formation. *Geobiology*, **1**(1), 71-79.

Vandenbroucke, M., & Largeau, C. (2007). Kerogen origin, evolution and structure. *Organic Geochemistry*, **38**(5), 719-833.

Ventura, G. T., Kenig, F., Reddy, C. M., Schieber, J., Frysinger, G. S., Nelson, R. K., Dinel, E., Gaines, R. B., & Schaeffer, P. (2007). Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *Proceedings of the National Academy of Sciences*, **104**(36), 14260-14265.

Vermaas, W. F. (2001). Photosynthesis and Respiration in Cyanobacteria. *eLS*.

Vilchez, R., Pozo, C., Gómez, M. A., Rodelas, B., & González-López, J. (2007). Dominance of sphingomonads in a copper-exposed biofilm community for groundwater treatment. *Microbiology*, **153**(2), 325-337.

Von Der Weid, I., Korenblum, E., Jurelevicius, D., Rosado, A. S., Dino, R., Sebastian, G. V., & Seldin, L. (2008). Molecular diversity of bacterial communities from subseafloor rock samples in a deep-water production basin in Brazil. *Journal of microbiology and biotechnology*, **18**(1), 5-14.

Vossenberg, J. L., Ubbink-Kok, T., Elferink, M. G., Driessen, A. J., & Konings, W. N. (1995). Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea. *Molecular microbiology*, **18**(5), 925-932.

Vossenberg JLCM, Driessen AJM & Konings WN (2000) in *Cell and Molecular Response to Stress* (eds. Storey KB & Storey JM) 71-88. Elsevier.

Wächtershäuser, G. (1988). Before enzymes and templates: theory of surface metabolism. *Microbiological reviews*, **52**(4), 452.

Wächtershäuser, G. (1990). The case for the chemoautotrophic origin of life in an iron-sulfur world. *Origins of Life and Evolution of the Biosphere*, **20**(2), 173-176.

Wächtershäuser, G. (2000). Life as we don't know it. *Science*, **289**(5483), 1307-1308.

Wang, Y., Wiatrowski, H. A., John, R., Lin, C. C., Young, L. Y., Kerkhof, L. J., Yee, N., & Barkay, T. (2013). Impact of mercury on denitrification and denitrifying microbial communities in nitrate enrichments of subsurface sediments. *Biodegradation*, **24**(1), 33-46.

Webster, G., Blazejak, A., Cragg, B. A., Schippers, A., Sass, H., Rinna, J., Tang, X., Mathes, F., Ferdelman, T. G., Fry, J. C., Weightman, A. J., & Parkes, R. J. (2009). Subsurface microbiology and biogeochemistry of a deep, cold-water carbonate mound from the Porcupine Seabight (IODP Expedition 307). *Environmental microbiology*, **11**(1), 239-257.

Weiss, B. P., Yung, Y. L., & Nealson, K. H. (2000). Atmospheric energy for subsurface life on Mars?. *Proceedings of the National Academy of Sciences*, **97**(4), 1395-1399.

Wellsbury, P., Goodman, K., Barth, T., Cragg, B. A., Barnes, S. P., & Parkes, R. J. (1997). Deep marine biosphere fuelled by increasing organic matter availability during burial and heating. *Nature*, **388**(6642), 573-576.

Weerasooriya, R., & Dharmasena, B. (2001). Pyrite-assisted degradation of trichloroethene (TCE). *Chemosphere*, **42**(4), 389-396.

Wheat, C. G., McManus, J., Mottl, M. J., & Giambalvo, E. (2003). Oceanic phosphorus imbalance: Magnitude of the mid-ocean ridge flank hydrothermal sink. *Geophysical research letters*, **30**(17).

White, D. C., Stair, J. O., & Ringelberg, D. B. (1996). Quantitative comparisons of in situ microbial biodiversity by signature biomarker analysis. *Journal of Industrial Microbiology*, 17(3-4), 185-196.

White, D. C., & Ringelberg, D. B. (1997). Utility of the signature lipid biomarker analysis in determining the in situ viable biomass, community structure, and nutritional/physiologic status of deep subsurface microbiota. *The Microbiology of the terrestrial deep subsurface*, 119-136.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, **95**(12), 6578-6583.

Widerlund, A., & Davison, W. (2007). Size and density distribution of sulfide-producing microniches in lake sediments. *Environmental science & technology*, **41**(23), 8044-8049.

Wilcoxen, J., Zhang, B., & Hille, R. (2011). Reaction of the molybdenum-and copper-containing carbon monoxide dehydrogenase from Oligotropha carboxydovorans with quinones. *Biochemistry*, **50**(11), 1910-1916.

Wilkins, M. J., Daly, R. A., Mouser, P. J., Trexler, R., Sharma, S., Cole, D. R., Wrighton, K. C., Biddle, J. F., Denis, E. H., Fredrickson, J. K., Kieft, T. L., Onstott, T. C., Peterson, L., Pfiffner, S. M., Phelps, T. J., & Schrenk, M. O. (2014). Trends and future challenges in sampling the deep terrestrial biosphere. *Frontiers in microbiology*, **5**, 481.

Williams, J. P., & Hallsworth, J. E. (2009). Limits of life in hostile environments: no barriers to biosphere function?. *Environmental microbiology*, **11**(12), 3292-3308.

Wilmes, P., Andersson, A. F., Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L., Bond, P. L., VerBerkmoes, N. C., & Banfield, J. F. (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME journal*, **2**(8), 853-864.

Winfrey, M. R., & Zeikus, J. G. (1977). Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments. *Applied and Environmental Microbiology*, **33**(2), 275-281.

Woese, C. R. (1979). A proposal concerning the origin of life on the planet earth. *Journal of Molecular Evolution*, **13**(2), 95-101.

Wolfaardt, G. M., Lawrence, J. R., & Korber, D. R. (1999). Function of EPS. In *Microbial extracellular polymeric substances* (pp. 171-200). Springer Berlin Heidelberg.

Wolfaardt, G.M., Korber, D., Lawrence, J. R. (2007). Cultivation of microbial consortia and communities. In *Manual of environmental microbiology*, (Eds. Hurst, C. J., Crawford, R. L., Garland, J. L., Lipson, D. A., Mills, A. L., Stetzenbach, L. D.), 101-111.

Wolkers, W. F., Tablin, F., & Crowe, J. H. (2002). From anhydrobiosis to freeze-drying of eukaryotic cells. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **131**(3), 535-543.

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., Malmstrom, R., Stepanauskas, R., & Cheng, J. F. (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One*, **6**(10), e26161.

Yamazaki, S. I., Kaneko, T., Taketomo, N., Kano, K., & Ikeda, T. (2002). Glucose metabolism of lactic acid bacteria changed by quinone-mediated extracellular electron transfer. *Bioscience, biotechnology, and biochemistry*, **66**(10), 2100-2106.

Yarwood, C. E. (1950). Water content of fungus spores. *American Journal of Botany*, **37**, 636-639.

Yatsusenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M.,

Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R., & Gordon, J. (2012). Human gut microbiome viewed across age and geography. *Nature,* **486**(7402), 222-227.

Yilmaz, S., Allgaier, M., & Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature methods*, **7**(12), 943-944.

Yoon, S. S., Kim, E. K., & Lee, W. J. (2015). Functional genomic and metagenomic approaches to understanding gut microbiota–animal mutualism. *Current opinion in microbiology*, **24**, 38-46.

Zambrano, M. M., Siegele, D. A., Almiron, M., Tormo, A., & Kolter, R. (1993). Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science*, **259**(5102), 1757-1760.

Zhang, L., Keller, J., & Yuan, Z. (2009). Inhibition of sulfate-reducing and methanogenic activities of anaerobic sewer biofilms by ferric iron dosing. *Water research*, **43**(17), 4123-4132.

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and environmental microbiology*, **62**(2), 316-322.

Zhou, J., Liu, S., Xia, B., Zhang, C., Palumbo, A. V., & Phelps, T. J. (2001). Molecular characterisation and diversity of thermophilic iron-reducing enrichment cultures from deep subsurface environments. *Journal of applied microbiology*, **90**(1), 96-105.

Zou, Y., Pisciotta, J., Billmyre, R. B., & Baskakov, I. V. (2009). Photosynthetic microbial fuel cells with positive light response. *Biotechnology and bioengineering*, **104**(5), 939-946.

ZoBell, C. E., & Anderson, D. Q. (1936). Vertical distribution of bacteria in marine sediments. *AAPG Bulletin*, **20**(3), 258-269.

ZoBell, C. E. (1938). Studies on the bacterial flora of marine bottom sediments. *Journal of Sedimentary Research*, **8**(1).

ZoBell, C. E., & Morita, R. Y. (1957). Barophilic bacteria in some deep sea sediments. *Journal of bacteriology*, **73**(4), 563.

12. PUBLICATIONS

Iberian Pyrite Belt Subsurface Life (IPBSL), a drilling project of biohydrometallurgical interest

Advanced Materials Research Vol. 825 (2013) pp 15-18
© (2013) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMR.825.15

Iberian Pyrite Belt Subsurface Life (IPBSL), a drilling project of biohydrometallurgical interest

Ricardo Amils[1, 2, a *], David Fernández-Remolar[1, b], Victor Parro[1, c], José Antonio Rodríguez-Manfredi[1,d], Ken Timmis[3,e], Mónica Oggerin[1, f], Mónica Sánchez-Román[1, g], Francisco, J. López[1, h], José Pablo Fernández[1, i], Fernando Puente[1, j], David Gómez-Ortiz[4, k], Carlos Briones[1, l], Felipe Gómez[1, m], Enoma Omoregie[1, n], Miriam García[1, o], Nuria Rodríguez[1,p], José Luis Sanz[5, q] and the IPBSL Team

[1]Centro de Astrobiología (INTA-CSIC), 28850 Torrejón de Ardoz, Madrid, Spain.

[2]Centro de Biología Molecular Severo Ochoa (UAM-CSIC), 28049 Madrid, Spain.

[3]Technical University of Braunschweig, Germany

[4]ESCET-Area de Geología, Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain.

[5] Departamento de Biología Molecular, UAM, Cantoblanco 28049 Madrid, Spain.

[a]ramils@cbm.uam.es, [b]fernadezrd@cab.inta-csic.es, [c]parrogv@cab.inta-csic.es,

[d]manfredi@cab.inta-csic.es, [e]emi.mbt@gouglemail.com, [f]oggerinom@cab.inta-csic.es,

[g]msanzroman@cab.inta-csic.es, [h]lopezsfj@cab.inta-csic.es, [i]fernandezrjp@cab.inta-csic.es,

[j]puentesf@cab.inta-csic.es, [k]david.gomez@urjc.es, [l]cbriones@cab.inta-csic.es,

[m]gomezgf@cab.inta-csic.es, [n]omoregie@cab.inta-csic.es, [o]villadangosgm@cab.inta-csic.es,

[p]nrodriguez@cbm.uam.es, [q]joseluis.sanz@uam.es

Keywords: Iberian Pyrite Belt, subsurface geomicrobiology, iron cycle, sulfur cycle, methanogenesis, Sulfate Reducing Bacteria (SRB)

Abstract. The geomicrobiological characterization of Río Tinto, an extreme acidic environment, has proven the importance of the iron cycle, not only in generating the extreme conditions of the habitat (low pH, high concentration of toxic heavy metals) but also in maintaining the high level of microbial diversity detected in the water column and the sediments. The extreme conditions detected in the Tinto basin are not the product of industrial contamination but the consequence of the presence of an underground bioreactor that obtains its energy from the massive sulfide minerals of the Iberian Pyrite Belt (IPB). To test this hypothesis, a drilling project (IPBSL) to intersect ground waters interacting with the mineral ore is under way, to provide evidence of subsurface microbial activities. A dedicated geophysical characterization of the area selected two drilling sites due to the possible existence of water with high ionic content. Two wells have been drilled in Peña de Hierro, BH11 and BH10, with depths of 340 and 620 meters respectively, with recovery of cores and generation of samples in anaerobic and sterile conditions. The geological analysis of the retrieved cores showed an important alteration of mineral structures associated with the presence of water, with production of expected products from the bacterial oxidation of pyrite. Ion chromatography of water soluble compounds from uncontaminated samples showed the existence of putative electron donors, electron acceptors, as well as variable concentration of metabolic organic acids, which suggest the presence of an active subsurface ecosystem associated to the high sulfidic mineral content of the IPB. Enrichment cultures from selected samples showed evidences of an active iron and sulfur cycle, together with unexpected methanogenic, methanotrophic and acetogenic activities. The geological, geomicrobiological and molecular biology analyses which are under way, should allow the characterization of this ecosystem of biohydrometallurgical interest.

Introduction

Río Tinto is an unusual ecosystem due to its size, constant pH, high concentration of heavy metals and high level of microbial diversity [1]. Río Tinto rises in Peña de Hierro, in the core of the Iberian Pyrite Belt (IPB). The IPB is one of the largest massive sulfide deposits on Earth. One important characteristic of Río Tinto is the high concentration of ferric iron and sulfates present in its waters, products of the biooxidation of pyrite, the main mineral component of the IPB. The IPBSL project was designed to answer basic questions related with the subsurface geomicrobiology responsible of the extreme conditions detected in the Tinto basin (http://auditore.cab.inta-csic.es/ipbsl).

Methodology

Boreholes were continuously cored by rotary diamond-bit drilling using a Boart-Longyear HQ wireline system producing 3 meters of 60 mm diameter cores. Well water was used as drilling fluid to lubricate the bit and return cuttings to the surface. Fluids were re-circulated. To detect potential contamination of the samples, sodium bromide (200 ppm) was added to the drilling fluid as a marker. Upon retrieval from the drilling rig, cores were divided into 60 cm length pieces, inspected for signs of alteration and stored in boxes for its permanent storage and curation in the Instituto Geológico Minero de España (IGME) lithoteque in Peñaroya. Selected cores were deposited in plastic bags, oxygen was displaced with N_2, sealed and transported to a field laboratory within 60 minutes. After drilling, boreholes were cased with PVC tubes with holes at different depths to allow water movement. Upon arrival at the field laboratory cores were placed in an anaerobic chamber (5% H_2, 95% N_2), logged and photographed. Aseptic subsamples were obtained by splitting cores with an hydraulic core splitter and drilling out the central untouched portion with a rotary hammer with sterile bits, with strict temperature control (40°C maximum). Rock leachates were produced by adding 5 ml sterile water to 0.5 g of powdered core subsamples and allowing them to stand overnight before filtration through pre-rinsed nylon 0.2 μm filters and analysed in a Advanced Compact Ion Chromatographer IC (Metrohm AG). XRD analysis was done with a Seifert 3003 T-T X-Ray difractometer. Elemental analysis was done by ICP-MS using a ELAN-6000 PE-Sciex instrument. Gases were analysed by gas chromatography using a Shimadzu GC-8A equipped with a 2 m glass column packed with Poroàck Q. Samples for DNA extraction were kept at 4°C. Samples for RNA were mixed with 2.5 volumes of LifeGuard to preserve and increase the stability of the RNA molecules. Previous test in our lab have demonstrated that the addition of EDTA increases nucleic acids stability in this type of samples. It has been described that pyrite can induce RNA degradation in few hours. Since pyrite is the main component present in the IPB subsurface, 100 mM EDTA was added to the LifeGuard reagent. We also avoid freeze the samples to prevent cell lyses. Environmental DNA was extracted using the commercial MoBio DNA extraction kit from soil. Cloning into plasmid vectors, sequencing and phylogenetic studies were performed as previously described [2, 3]. Powder and chips from different core samples were directly analysed by Sandwich Microarray Immunoassay (SMI) as described elsewhere [3]. Culture independent detection of microorganisms was done by epifluorescent microscopy after staining samples with DAPI and hybridization with universal probes for Bacteria (EUB388) and Archaea (Arch915) using CARD-FISHS. Samples were fixed with 4% formaldehyde and stored at -20°C in ethanol:PBS (1:1) until further processing. Fluorescence in situ hybridizations were done following the protocols described in [2]. Chemolithotrophic enrichment cultures were performed in a minimal Mackintosh medium [4], with the addition of ferrous iron or a sterile pyrite sample as electron donors. Anaerobic enrichments for denitrifying microorganisms were performed as described by Stevens and McKinley [5], for sulfate reducers according to González.-Toril et al. [2], and for methanogens according to Sanz et al. [6].

Results

Two distinctive geological activities were performed concerning the selection of the drilling sites. Firstly, a careful mapping and surface sampling for solid rocks and springs in the Peña de Hierro area was performed to evaluate all available geological information. Secondly, different geophysical procedures were applied to detect the most probable subsurface areas hosting microbial activity in deep regions of the basement (Transient Electromagnetic sounding and 1D resistivity logging). These measurements were followed by emplacing two different resistivity lines that allowed determining the structure and lithological distribution of the sites with high interest for drilling. From the analysis of the geophysical information it was decided to drill two wells, BH10 and BH11, with depths of 620 and 340 meters respectively.

In addition to the geological core login in the drilling site, selected samples were obtained for mineralogical (XRD), elemental analysis (ICP-MS), and stable isotopic and petrographic analysis. The mineralogical results showed the presence of pyrite and alteration products like hematite and magnesite in both boreholes. The elemental analysis of leachates from the solid samples showed the presence of iron and other metals at different depths in both boreholes, indicating an alteration of the metal sufides at specific positions along the boreholes. The stable isotopic analysis of pyrites showed important sulfur fractionation at different depths in samples from both boreholes, which is a clear indication of activity of sulphate reducing microorganisms at these depths, which have been corroborated by enrichment cultures (see below). Also carbon fractionation signals were obtained in samples from both boreholes, being also a clear biosignature of microbial activity at these depths.

Rock leachates were analysed by ion chromatography to determine the concentration of water soluble anions. The obtained results indicate the presence of reduced organic anions like acetate, formate and propionate and oxidized inorganic anions such as nitrate, nitrite and sulfate. The appearance of these compounds at different depths is a strong indication of biological activity along the boreholes. Total protein and sugar content was also detected at different depths, indicating the presence of extant or recent microbiological activity. Samples along the entire length of borehole BH10 were analysed with the immunosensor LDChip450, an antibody microarray containing 450 antibodies against microbial cells, environmental extracts, proteins, exopolysaccharides, etc. More than 40 core samples were analysed and plotted to compare the immunorpofiles at different depths. It was observed high biomarker detection at -392 m and around the interval between -500 and -550 m in borehole BH10. Positive Ag-Ab reactions were detected with specific antibodies against methanogenic archaea and SRB which agree with the results obtained with enrichment cultures (see below). H_2, CO_2 and CH_4 have been detected in mineral samples from both boreholes.

DNA and RNA have been successfully extracted from different BH10 and BH11 samples. Most of them rendered positive PCR amplifications of the bacterial 16S rRNA gene.

During the drilling campaign samples from the two boreholes were collected (47 samples for BH10 and 21 for BH11) for fluorescence in situ hybridization. The results obtained so far showed positive hybridizations at different depths (-207, -352, -497 and -608 meters) for borehole BH10. The detected microorganisms were grouped in colonies attached to mineral particles, thus is unlikely that they might correspond to contamination, especially because all the selected samples had the background Br concentration (less than 0.5%). Further hybridization with specific probes selected or designed after identification of putative organisms using 16S rRNA gene sequences of the same samples is under development.

Anaerobic enrichment cultures have been prepared in the anaerobic chamber using mineral salts medium with the addition of different electron donors (ferrous iron, pyrite, thiosulfate, H_2, acetate, mixture of organic acids) and electron acceptors (ferric iron, sulfate, nitrate, CO_2, arsenate). The

following activities have been detected unambiguously after more than 10 months of incubation: methanogens, methanotrphs, sulfur reducers, iron oxidizers, acetogens and denitrifiers, using samples from both boreholes. From all the available data two hot spots have been detected in BH10, one at -352 m and at -497 m, and two in BH11, one at –236 m and at -311 m. The identification of hotspots is required for the selection of samples for metagenomic and retrotranscriptomic analysis, which is under way.

Conclusions

The results obtained so far allowed to reach the following conclusions. As groundwater enters the Volcanogenic-hosted Massive Sulfide (VHMS) system, biological and abiotc processes are activated. Electron acceptors available for microbial metabolism include transient oxygen, nitrate, sulfate, ferric iron and inorganic carbon. Electron donors include ferrous iron, sulfide, H_2 generated by water/rock interaction, supporting the generation of methane and organic acids, like acetate. This supports a community of different microbial metabolisms. As the fluid becomes more reduced, methanogenesis and sulfate reduction, using hydrogen, become the dominant microbial processes. Oxidants to drive the system appear to be supplied by the rock matrix, in contrast to conventional ARD models. Only mobilization of these sources by ground water appears to be necessary to allow microbial metabolism. These observations confirmed the hypothesis that microorganisms are active in the subsurface of the IPB. The characterization of these activities is extremely important to gain inside on the microbial ecology that might be operating in heap leaching processes, in some cases affecting negatively their efficiencies.

References

[1] E. Gonzalez-Toril, E. Llobet-Brossa, EO. Casamayor, R. Amann, R. Amils, (2003) Microbial ecology of an extreme acidic environment, the Tinto River. Appl Environ Microbiol 69: 4853-4865.

[2] E. Gonzáklez-Toril, F. Gómez, M. Malki, R. (2006) Isolation and study of acidophilic microorganisms. In "Methods in Microbiology", F. Rainey and A. Oren (eds.), Elsevier, Oxford, Vol. 35, pp. 463-502.

[3] V. Parro (2010) Antibody microarrays for environmental monitoring. In "Handbook of Hydrocarbon and Lipid Microbiology", K.N. Timmis (ed.), Springer-Verlag, Berlin, pp. 2699-2710.

[4] M.E. Mackintosh (1978). Nitrogen fixation by Thiobacillus ferrooxidans. J. Gen. Microbiol., 105: 215-218.

[5] T. Stevens, J.P. McKinley (1995), Lithoautotrophic microbial ecosystems in deep basalt aquifers, Science, 250: 450-454..

[6] J.L. Sanz, N. Rodriguez, R. Amils (1997). "Effect of chlorinated aliphatic hydrocarbons of the acetoclastic methanogenic activity of granular sludge", Appl. Microbiol. Biotechnol., 47, 324-328.

Deep subsurface sulfate reduction and methanogenesis in the Iberian Pyrite Belt revealed through geochemistry and molecular biomarkers

Geobiology

Geobiology (2014), **12**, 34–47 DOI: 10.1111/gbi.12065

Deep subsurface sulfate reduction and methanogenesis in the Iberian Pyrite Belt revealed through geochemistry and molecular biomarkers

F. PUENTE-SÁNCHEZ,[1] M. MORENO-PAZ,[1] L. A. RIVAS,[1] P. CRUZ-GIL,[1] M. GARCÍA-VILLADANGOS,[1] M. J. GÓMEZ,[1] M. POSTIGO,[1] P. GARRIDO,[1] E. GONZÁLEZ-TORIL,[2] C. BRIONES,[1] D. FERNÁNDEZ-REMOLAR,[2] C. STOKER,[3] R. AMILS[2,4] AND V. PARRO[1]

[1]*Departments of Molecular Evolution, Centro de Astrobiología (INTA-CSIC), Madrid, Spain*
[2]*Planetology and Habitability, Centro de Astrobiología (INTA-CSIC), Madrid, Spain*
[3]*NASA Ames Research Center, Moffett Field, CA, USA*
[4]*Centro de Biología Molecular "Severo Ochoa" (UAM-CSIC), Cantoblanco, Madrid, Spain*

ABSTRACT

The Iberian Pyrite Belt (IPB, southwest of Spain), the largest known massive sulfide deposit, fuels a rich chemolithotrophic microbial community in the Río Tinto area. However, the geomicrobiology of its deep subsurface is still unexplored. Herein, we report on the geochemistry and prokaryotic diversity in the subsurface (down to a depth of 166 m) of the Iberian Pyritic belt using an array of geochemical and complementary molecular ecology techniques. Using an antibody microarray, we detected polymeric biomarkers (lipoteichoic acids and peptidoglycan) from Gram-positive bacteria throughout the borehole. DNA microarray hybridization confirmed the presence of members of methane oxidizers, sulfate-reducers, metal and sulfur oxidizers, and methanogenic Euryarchaeota. DNA sequences from denitrifying and hydrogenotrophic bacteria were also identified. FISH hybridization revealed live bacterial clusters associated with microniches on mineral surfaces. These results, together with measures of the geochemical parameters in the borehole, allowed us to create a preliminary scheme of the biogeochemical processes that could be operating in the deep subsurface of the Iberian Pyrite Belt, including microbial metabolisms such as sulfate reduction, methanogenesis and anaerobic methane oxidation.

Received 24 July 2013; accepted 8 October 2013

Corresponding author: V. Parro. Tel.: +34-915201071; fax: +34-915201074; e-mail: parrogv@cab.inta-csic.es

INTRODUCTION

The geomicrobiology of subsurface environments is a matter of growing interest not only for the understanding of life in the absence of light and its role in biogeochemical cycles, but also as a model for searching for life in other planetary bodies where the surface conditions preclude any form of life. The Río Tinto fluvial basin (southwestern Spain) is an acidic system in continuous geological evolution dating back from more than 2 million years (Fernández-Remolar *et al.*, 2005; Fernández-Remolar & Knoll, 2008). It emerges in the Rio Tinto Anticline region, which is a complex geological structure that is part of the Iberian Pyrite Belt (IPB) (Leistel *et al.*, 1997). Its formation is associated with the building up of a Carboniferous volcano-sedimentary complex exposed to a late intensive event of hydrothermal activity that mineralized the volcanic systems. Superficial and subsurface microbial communities sustained on sulfur and iron chemolithotrophy (Amils *et al.*, 2007) have been using the ancient hydrothermal materials as energy sources for millions of years. The microbial activity accelerates the oxidation of the pyrite ore body with the subsequent release of protons, which account for the extreme acidic conditions (pH <3) observed in the modern fluvial system. Additionally, recent studies have shown that some regions of the Río Tinto subsurface habitats have neutral pH (Fernández-Remolar *et al.*, 2008b) and that the presence of carbonate minerals similar to those found in ancient martian rocks

could be a consequence of microbial activity (Fernández-Remolar *et al.*, 2012).

During the Mars Astrobiology Research and Technology Experiment (MARTE) project (Stoker *et al.*, 2008), four boreholes (BH) in the iron-sulfide deposits at Peña de Hierro, an abandoned mine site located in the catchment area of the Río Tinto, were performed: BH1, BH4, BH8, and BH7 with depths of 59, 166.35, 166, and 5 m, respectively. The objectives were the development of technology for *in situ* drilling and life detection in planetary exploration, as well as a preliminary geochemical and geomicrobiological characterization of the subsurface. It was a project of marked astrobiological character with a strong technological component, with drilling, aseptic sampling, and searching for microbial activities in a sulfide deposit as the main challenges. Previous published studies from MARTE were focused on the drilling, life detection technology development (Stoker *et al.*, 2008; Parro *et al.*, 2008), and geochemical and mineralogical analyses (Fernández-Remolar *et al.*, 2008a,b). In Parro *et al.* (2008), we reported the results of a field testing of the SOLID2 instrument for the analysis of BH7 (shallow drill up to 5 m). The objective was instrument validation, and very limited biological information was presented.

In the present work, we focus on the anaerobic part of the BH8 (below ca. 90 m below the surface of the IPB) and present novel and substantial information about: (i) the geochemistry, particularly the determination of organic acids (acetate, formate, oxalate) and nitrate, nitrite, and sulfate anions; (ii) the detection of complex organic matter and microbial biopolymers; and (iii) the microbial community.

The application of molecular ecology techniques is critical to elucidate the geomicrobiology of deep subsurface environments. The heterogeneity (Haldeman *et al.*, 1993) of mineral subsurface samples and the inherent bias of each technique make the combination of several complementary or even redundant approaches especially useful for this purpose. In a previous work (Rivas *et al.*, 2008), we described an antibody microarray-based biosensor containing more than 200 antibodies against bacterial and archaeal strains, crude natural extracts, cellular proteins and polysaccharides, and other biomolecules. We reported the usefulness of the EMCHIP200 (for Environmental Monitoring CHIP) for immunoprofiling of environmental samples. Additionally, we tested an updated version of the chip (LDChip300, for *Life Detector Chip* with 300 antibodies) for subsurface life exploration during different drilling campaigns (Parro *et al.*, 2011; Blanco *et al.*, 2012). On the other hand, we demonstrated the usefulness of a small oligonucleotide microarray (the PAM microarray) to explore the prokaryotic diversity at various challenging environments (Garrido *et al.*, 2008; Parro *et al.*, 2011; Blanco *et al.*, 2012). In the present work, the complementary analysis of core samples with multiple molecular ecological tools, such as antibody microarrays (e. g. LDChip), oligonucleotide arrays (PAM microarray), and sequencing, allowed us to infer the relationship between the geochemistry, the prokaryotic diversity, and the potential metabolisms in this deep subsurface environment.

MATERIALS AND METHODS

Drilling and sampling

During the 2004 field campaign of the MARTE project, a joint project between NASA Ames Research Center (funded by NASA's ASTEP program) and the Centro de Astrobiología (INTA-CSIC), a 164-m-depth borehole was created using a wireline drilling system (Fernández-Remolar *et al.*, 2008b). Coring was performed with a Boart-Longyear (Salt Lake City, UT, USA) HQ wireline system that produced 60-mm-diameter cores within a plastic liner. A chemical tracer (NaBr) was incorporated into the drilling fluid to identify possible sample contamination produced during drilling by using ion chromatographic analysis of aliquots of the samples analyzed for biology. Cores were retrieved in 1-m sections, encased in plastic liners, placed in sleeves filled with N_2 gas at the drilling site, sealed, and transported to a field laboratory located at the Museo Geominero in the Ríotinto village. In the laboratory, cores were placed and sampled inside a N_2-filled anaerobic chamber as described by Fernández-Remolar *et al.* (2008b). Briefly, powder samples were extracted from the center of core pieces that were broken open using a core cutter. Samples of rock cuttings were acquired in each 1-m section using a hand-drill and a sterilized bit. Sampling locations were chosen with preference for any parts of the cores that showed evidence of alteration (e.g. oxidation). All elements in contact with rock cores were cleaned and sterilized prior to use on every new sample. Thus obtained, powdered sample was distributed in different sets of tubes as a function of the analysis to be performed, and, once out of the anaerobic chamber, the samples for microbiological and molecular studies were stored frozen (−20°C) or immediately processed. To minimize the effects of contamination by the drilling fluid, samples in which bromide was detected by ion chromatography were excluded from further analysis.

The aim of this work was the geomicrobiological study of the anaerobic subsurface (below ca. 90 m) of the IPB. Thus, we focused this work on the study of the geomicrobiology associated with the microaerobic and the anaerobic zones of BH8, below the water table, which in BH8 occurred at 90 m depth (Fig. 1; for a geological map of the area, see Fernández-Remolar *et al.*, 2008b; Fig. 1).

Ion chromatography analysis

Core samples were analyzed for the concentrations of inorganic anions such as nitrite, nitrate, and sulfate, and small-molecular-weight organic acids such as acetate, for-

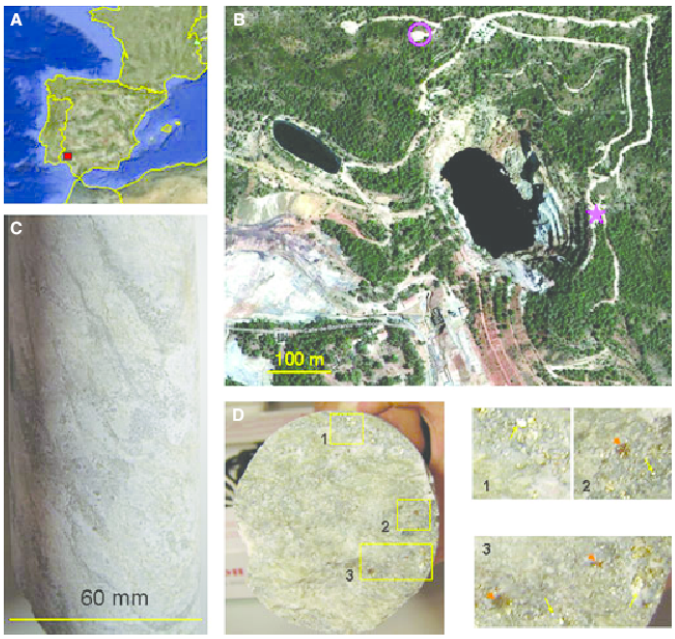


Fig. 1 Exploration of the Iberian Pyrite Belt (IPB) subsurface geomicrobiology. Geographical location of the IPB in the Iberian Peninsula (A) and a satellite view (B) of Peña de Hierro (Nerva, Spain), where the 2004 MARTE project drilling campaign was carried out (star). Another drilling site currently under study is highlighted (circle). Note the gray/light blue material at the side walls of the lake, which corresponds to the massive pyrite stock work. The vertical distance between the drilling site and the lake water table was about 90 m. (C) Profile of one of the cores recovered from 130 mbs. (D) Cross section of the same core. Numbered squares (amplified on the right) indicate interesting features in the core, such as oxidized red spots (arrowheads) or the high abundance of bright pyrite crystals (small arrows).

mate, and oxalate using ion chromatography as described (Parro *et al.*, 2011). Briefly, 10 g of sample were sonicated (3 × 1 min cycles) in 20 mL of water, and the mineral particles were removed by centrifugation. Supernatants were collected and loaded into a Metrohm 861 Advanced Compact Ion Chromatographer (Metrohm AG, Herisau, Switzerland) undiluted, or at dilutions values of either 50% or 20%, using a *Metrosep A supp 7-250* column, with 3.6 mM sodium carbonate ($NaCO_3$) as eluent.

Each sample was measured at three different dilutions to ensure that concentration ranges were within the calibration curve for each anion (see below). The measurement error of the equipment for replicate samples was <1%. The concentration ranges for each calibration curve were the following: fluoride, bromide, propionate, formate, phosphate, tartrate, and oxalate from 0.08 (lower limit) to 2 ppm (higher limit); acetate, from 0.2 to 5 ppm; chloride, from 0.4 to 10 ppm; nitrite, from 0.2 to 5 ppm; nitrate from 0.2 to 10 ppm; and sulfate, from 8 to 200 ppm.

Whole crude extract from cores and chemical analysis

Extracts were prepared from 40 g of a sample obtained at 139.38 m depth with GuHCl buffer (4 M guanidine-hydrochloride, 0.5 M EDTA, 0.5 M Tris, pH 7.4) following the procedure described by Tuross & Stathoplos (1993). The final extracts were re-suspended in sterile distilled water, dialyzed (>1200 Da cutoff) against water, and finally lyophilized. These extracts were used for thin-layer chromatography (TLC) fractionation to separate peptidoglycan fragments and lipopolysaccharides (LPS) containing compounds. TLC was run on silica-gel 60 F_{254} TLC plates (Merck) using isobutyric acid: 1N ammonium hydroxide: water (5:3:2 v/v/v) as solvent, and the detection was performed by short-wave UV fluorescence and by vanillin staining (Rivas *et al.*, 2008). The spots stained with vanillin were scrapped from the plate, eluted in distilled water, and analyzed by mass spectrometry using electrospray ionization-tandem mass spectrometry (ESI-MS/MS) in the Laboratorio de Espectrometría de Masas (SIDI, Universidad Autónoma de Madrid, Spain). Two mass spectrometers were used: an ion trap LCQ Deca XP plus (Thermo Fisher Scientific, Waltham, MA, USA) and a hybrid analyzer, quadrupole time-of-flight (QTOF) model QSTAR pulsar I (Applied Biosystems, Foster City, CA, USA). The molecular composition was obtained using *Elemental composition calculator* (Wsearch software, Melbourne, Victoria, Australia) and freely available public software (e.g. http://www.chemspider.com) was used for the estimation of the most probable chemical structure for a given molecular formula.

Antibodies: production, purification, fluorescent labeling, and microarray production

The antibodies used in this study were purified with protein-A and fluorescently labeled with Alexa-647

fluorochrome (Invitrogen, Carlsbad, CA, USA) as described previously (Rivas *et al.*, 2008). The 200 antibodies described by Rivas *et al.* (2008) were printed onto epoxy-activated microscope slides (Arrayit Corp., Sunnyvale, CA, USA) to form the so-called LDChip200. Printing was performed in such a way that nine identical LDChips were allocated per microscope slide, regularly spaced to match with a cassette containing the same number of hybridization chambers, the multiple-array analysis module (MAAM) device (Parro, 2010).

Sandwich microarray immunoassay (SMI) with LDChip200

Powder and shards from different core samples (0.5 g) were directly analyzed by SMI as described elsewhere (Parro, 2010; Parro *et al.*, 2011). Briefly, the potential molecular biomarkers or cells present in the sample were extracted in 1.5–2.0 mL of buffer TBSTRR (0.4 M Tris-HCl, pH 8, 0.3 M NaCl, 0.1% Tween 20) by means of a handheld ultrasonicator (DR. Hielscher 50W DRH-UP50H sonicator, Hielscher Ultrasonics, Berlin, Germany). The sample was then filtered through a 15-μm filter, and 40 μL of the filtrate was incubated in an LDChip200 containing 200 antibodies against microbial cells, environmental extracts biological polymers, such as lipoteichoic acids and peptidoglycan, proteins. (Table S1; Rivas *et al.*, 2008). Up to eight samples plus a control can be simultaneously assayed per microscope slide in a MAAM device. After a washing step to remove the non-specifically bound material and a second incubation with a mixture of 200 fluorescent tracer antibodies, the positive antigen–antibody reactions were read by exciting the fluorochrome at 635 nm and measuring the fluorescence emission at 650 nm in a commercial scanner (GenePix4100A, Genomic Solutions, Huntingdon, UK). The scanned images were analyzed and quantified using commercial microarray analysis software (Genepix™ Pro Software, Genomic Solutions) as described (Parro *et al.*, 2011; Blanco *et al.*, 2012).

DNA extraction and amplification

Environmental DNA was extracted from 2 g of powdered cores by using the commercially available MoBio DNA extraction kit for soil (MoBio Labs, Solana Beach, CA, USA). The DNA was used as template for simultaneous PCR amplification and fluorescent labeling using the universal primers for the prokaryotic 16S rRNA gene. The primers were 8F (positions 8–27 of *Escherichia coli* 16S rRNA, 5′-AGAGTTTGATCATGGCTCAG), and either 1492R (positions 1492–1474, 5′-GGTTACCTTGTTACGACTT) or 16SR (positions 1057–1074, 5′-CACGAGCTGACGACAGCCG). The thermocycler was programmed as follows: 95°C, 5 min; 10× (95 °C 20 s, 50 °C 30 s, 68 °C 60 s); 25× (95 °C 20 s, 48 °C 30 s, 68 °C 60 s + 5 s per cycle); 68°C 10 min; 4°C.

Because the quality and yield of the obtained DNA were very low, most of the samples were subjected to whole metagenome amplification by isothermal multiple displacement amplification (MDA) with phi29 DNA polymerase (GenomiPhi DNA Amplification Kit, GE Healthcare, Little Chalfont, Buckinghamshire, UK), formerly Amersham Bioscience). The amplified product was used for PCR amplification or directly labeled with a fluorophore and analyzed with the PAM oligonucleotide microarray (see below).

Oligonucleotide microarray hybridization

Garrido *et al.*, 2008 previously reported the so-called prokaryotic acidophile microarray (PAM), an oligonucleotide microarray containing specific 16S and 23S rRNA gene probes for prokaryotic acidophiles (see Garrido *et al.*, 2008; Table S1). Although PAM was specially developed for the detection of prokaryotic acidophiles in the extremely acidic environment of Río Tinto (Huelva, Spain), it also contains specific oligonucleotide probes for the 16S and 23S rRNA genes from other phylogenetic groups, such as α-, β-, γ-, and deltaproteobacteria, low-GC-content Firmicutes, high-GC-content Gram-positive bacteria, Archaea, Euryarchaeota, Chrenarchaeota, Methanobacteria, Eukarya. Fluorescently labeled 16S rRNA gene amplicons were checked by agarose gel electrophoresis, then purified with Qiagen PCR purification kit columns (Qiagen, CA, USA), and set to hybridize to the microarray in HibIt hybridization buffer (Arrayit Corp.) at 50°C for 12 h in a water bath. Then, they were washed and scanned for fluorescence (Garrido *et al.*, 2008). The specificity and the cross-hybridization of the probes in PAM microarray were tested by hybridization with the PCR-amplified 16S rRNA gene of the corresponding type strain (Garrido *et al.*, 2008).

Cloning sequencing and phylogenetic analysis

The extracted DNA either directly or after one round of MDA was used as a template for 16S rRNA gene amplification by PCR. Archaeal sequences were amplified using the archaeal-specific primer 20F (TTCCGGTTGATCCYGC CRG) paired with the universal primer U1392R (ACGGG CGGTGTGTRC), while bacterial sequences were amplified as described in the previous section. Cloning into plasmid vectors, sequencing, and phylogeny studies were performed as previously described (Parro *et al.*, 2011).

Fluorescent *in situ* hybridization (FISH) and scanning electron microscopy (SEM)

Samples of rock particles from the cores were gently washed to remove the fixative and the pore water. Drained

rock samples were immersed in 0.5% agarose to avoid detaching bacteria from the sample particles during incubation and washing. Hybridization and staining with DAPI (4′,6-diamidino-2-phenylindole) were carried out as previously described (Amann, 1995). Cy-3, Cy-5, and fluorescein isothiocyanate (FITC)-labeled probes for fluorescence *in situ* hybridization FISH were provided by Bonsai Technology (Barcelona, Spain). Catalyzed reporter deposition (CARD) was performed using the method previously described by Pernthaler *et al.*, 2002;. Horseradish peroxidase-labeled probes and the fluorochromes AlexaFluor488 and AlexaFluor534 were purchased from Molecular Probes (Invitrogen). Citifluor mounting medium (Citifluor Ltd, London, UK) was added to preparations to avoid fluorescence fading. An Axioskop microscope (Zeiss, Oberkochen, Germany) equipped with the proper filter set was used to visualize the FISH results. FISH and CARD-FISH results were further visualized under a fluorescence microscopy (Axioskop II, Zeiss) and LSM510 scanning confocal microscope (Zeiss) equipped with an Ar ion laser (458–514 nm) and two He/Ne lasers (543 and 633 nm). The probes used were EUB338 (GCT GCC TCC CGT AGG AGT), EUB338-II (GCA GCC ACC CGT AGG TGT), and EUB338-III (GCT GCC ACC CGT AGG TGT) (Amann, 1995 and Daims *et al.*, 1999), which are specific for bacteria, and ARCH915 (GTG CTC CCC CGC CAA TTC CT) (Stahl & Amann, 1991), which is specific for the Archaea domain. Scanning electron microscopy (SEM) of cores samples was performed as previously described (Fernández-Remolar *et al.*, 2008b).

RESULTS

Subsurface geochemistry

Core samples were analyzed by ion chromatography to identify the most relevant anions that could support microbial metabolisms (Fig. 2). Nitrate and nitrite were detected along the core length, the latter being more abundant with peaks of >45 μg g^{-1} at 94.7 and 156 m below the surface (mbs). High amounts of soluble sulfate were also measured, with maximal values of 18 mg g^{-1} at 94.7 and 156.4 mbs. Especially relevant was the detection of low-molecular-weight organic acids, such as acetate and formate, with maximum concentrations of >65.15 and > 15 μg g^{-1}, respectively, at 94.7 mbs. In addition, significant amounts of oxalate were also detected in one of the samples (156.4 mbs).

The concentration of soluble ferric iron (Fe^{3+}) was 20–50 μM in most of the samples, whereas that of ferrous iron (Fe^{2+}) showed a peak around 1.3 mM at 102 mbs. The Fe^{2+} concentration was very low from 156 mbs to the bottom of the borehole, an interval where most of the iron was in its oxidized state. Additionally, H_2 concentration was relatively constant at around 10 ppm in the anaerobic zone, while that of CH_4 increased from 142.5 mbs to the bottom of the borehole to a maximum of 31 ppm (V).

Detection of complex organic matter and biological polymers

Two crude extracts containing biological material (peptides or exopolymeric substances), one from a core containing altered pyrite at 139.38 mbs and the other from the surface of the drilling site, were analyzed using TLC with UV light and vanillin as the revealing agents. Dark spots were obtained with both detection techniques (Fig. 3A,B), revealing the presence of aromatic and steroid-like compounds. The spots from the 139.38 mbs sample were eluted from the TLC membrane and further analyzed by ion electrospray and time-of-flight (QTOF) spectrometry to identify their molecular composition and structure. The comparison of the assigned molecular formulae with those in the databases rendered a probable chemical structure for the main ions produced during the electrospray ionization procedure, which corresponded to aromatic, steroid, or triterpene-like compounds (Fig. 3, numbers 1–6).

Detection of microbial biomarkers using the antibody microarray LDChip200

Pulverized samples were processed by ultrasonication in an aqueous buffer and assayed with the LDChip200, an antibody microarray-based biosensor containing 200 antibodies against multiple biomolecules (Rivas *et al.*, 2008). The results indicated the presence of antigens recognized by the immobilized antibodies onto the microarray. These positive reactions corresponded to antibodies produced against antigens from Gram-positive bacteria (lipoteichoic acids, peptidoglycan), some steroid-like compounds, dsDNA, and biochemical extracts previously obtained from similar core samples (Rivas *et al.*, 2008). In addition, the LDChip200 detected ferritins and DPS-like proteins (Fig. 4) in some of the samples, as well as peptides related to L-asparaginase protein.

Prokaryotic diversity

Total environmental DNA was extracted to determine the microbial diversity by using molecular ecological techniques: (i) an oligonucleotide microarray specially designed for the detection of a broad group of prokaryote phyla (PAM microarray; Table S1; Materials and methods), and (ii) and sequencing of environmental 16S rRNA genes. The hybridization with 16S rRNA amplicons showed strong positive reactions with specific oligonucleotide probes for the Gram-positive low-GC-content Firmicutes phylum (LGC354a, b, c) (Fig. 5) in some of the samples (at 120.85 and 126.85 mbs) that previously

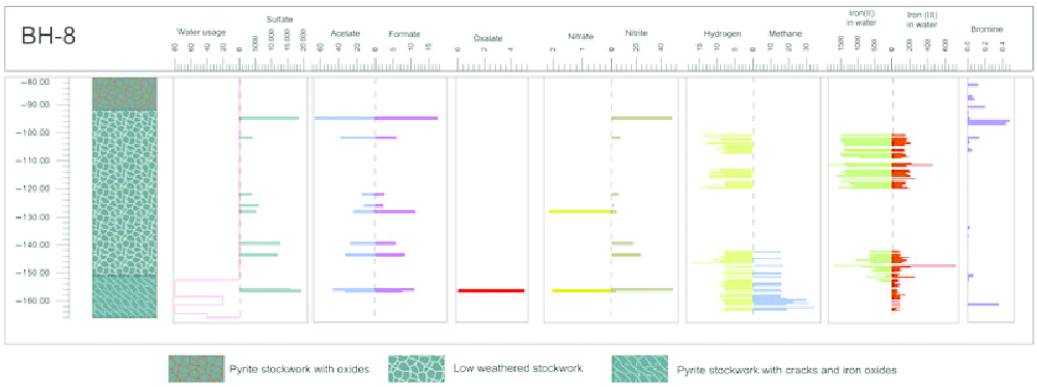


Fig. 2 Simplified lithostratigraphical section of borehole 8 showing lithology, water usage, acetate, formate, oxalate, nitrate, and bromine concentration in rock leachates. Hydrogen, methane, iron (II), and iron (III) concentrations were measured in solutions. Sodium bromide was used as a chemical tracer included in the drilling fluid, and the bromide concentration from the rock leachate was used to calculate the contamination factor. Water usage indicates the percentage of the drilling fluid loss through fractures during boring. All the concentration values are expressed in parts per million (ppm.).

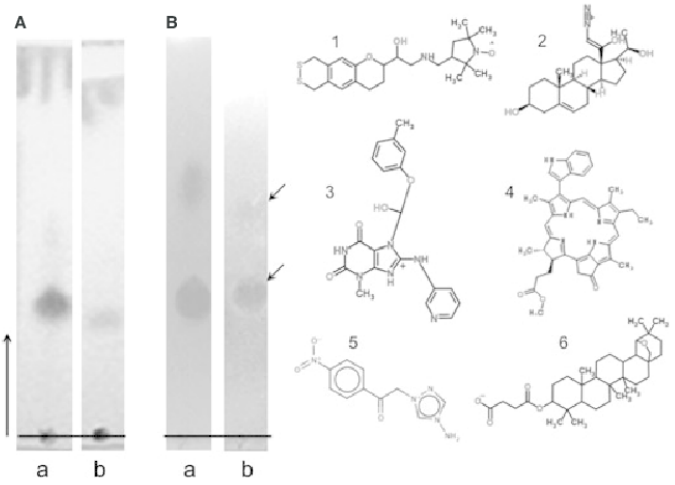


Fig. 3 Thin-layer chromatography of extracts from samples at the surface (a) and 139.38 m (b) depth. The result was revealed with short-wave UV light (A), indicating the presence of aromatic rings and conjugation, and with vanillin staining (B). Also shown are the most probable molecular structures obtained after the analysis of the 139.38 m spots (arrows) by ion electrospray (ESI+ and ESI-) and time-of-flight QTOF spectrometry: ***1-3***, ion 437.1938, which may correspond, respectively, to the molecular formula $C_{22}H_{33}N_2O_3S_2$, $C_{22}H_{33}N_2O_3$, or $C_{22}H_{25}N_6O_4$; ***4*** and ***6***, ion 637.3050, which may correspond, respectively, to the molecular formula $C_{40}H_{39}N_5O_3$ and $C_{34}H_{53}O_5$; ***5***, ion 343.9950, which may correspond to the molecular formula $C_{10}H_{10}N_5O_3$. More details can be found in Supporting Table S3.

exhibited a strong positive reaction with the anti-Gram-positive bacterial antibodies (Fig. 4). Among the members of the Firmicutes phylum that could contribute to these signals is the genus *Desulfotomaculum*, which gave a clear positive in the sample at 120.85 mbs (probe DFMI229). Other positive reactions corresponded to high-GC-content gram-positive bacteria (HGC236), the Actinobacteria genus *Ferrimicrobium* (probe FRM032), some *Acidithiobacillus* spp (*A. ferrooxidans* and *A. caldus*, probes ACT465a and THC642, respectively), *Thiomonas* spp group 2 (TM2G138), *Desulfovibrio* spp (DSV698), Euryarchaeota (ANME2c760), and Methanococcales (MC1109). The hybridization of the PAM microarray with the whole MDA-amplified metagenome confirmed

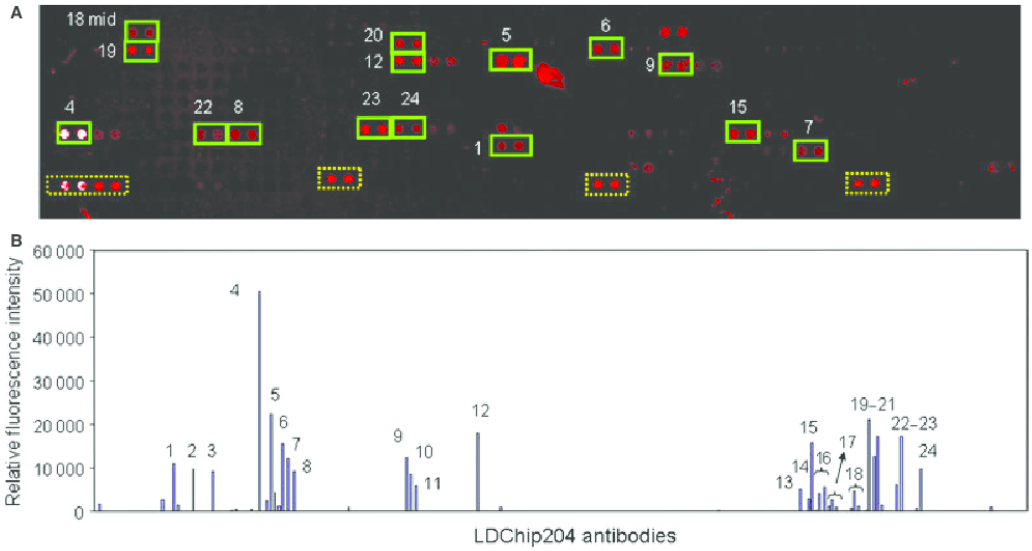


Fig. 4 Biomarker detection in deep IPB subsurface by the life detector chip LDChip200. A core sample (0.5 g) extracted at 150.15 mbs was analyzed with LDChip 200 by sandwich microarray immunoassay as indicated in Materials and methods. (A) The LDChip output is a 16-bits image with bright fluorescent spots (red to white, top picture) that showed positive antigen–antibody reactions. The fluorescent intensity can be quantified and plotted to form an immunogram (B). Due to the specificity of the antigen–antibody reaction, each positive signal in the chip immediately identifies the target compound: 1, steroid-like similar to estradiol; 2, ds/ssDNA; 3, *Cyanidium* spp rich natural extract; 4, Gram-positive antigen; 5, viral antigen; 6, dsDNA; 7, ferritin protein; 8, *Hydrogenobacter* spp-rich natural extract; 9, 10, 11, 12, antigens derived from extracts of IPB subsurface cores at 118–126 m, 138–139 m, 147–141 m, and 84–97 m deep, respectively (antibody names ID10S2, ID11S2, ID12S2, and ID7S2); 13, L-asparaginase protein; 14, GroEL protein; 15, bacterial Lipid A; 16, lipoteichoic acid (LTA) and *Listeria monocytogenes* LTA; 17, Lipid A, KDO bacterial antigen, and McrB protein; 18, NirS, NOR, and NRA; 19, penicillin derivatives; 20, peptidoglycan; 21, a pre-immune antiserum; 22, *Salmonella* spp; 23, cAMP; and 24, tryptophan.

the previous results and allowed the hybridization with other probes (Data S1). In particular, it revealed the presence of members of the Betaproteobacteria genus *Gallionella* (probe GALTS0084), the Deltaproteobacteria genus *Desulfosarcina/Desulfococcus* (DSS658), the WJ2-like bacterium, or the recurrent presence of high-GC-content gram-positive bacteria (probe HGC1901 from 23S rRNA gene). Among the Archaea, members of the Chrenarchaeota and Euryarchaeota divisions were found (MBGB380 and ANME2c760, respectively), as well as others belonging to the Methanosarcina family (MSSH8-59) and the Methanococcales order (MC1109).

Finally, PCR-amplified 16S rRNA gene from the samples which gave positive signals in the antibody microarray and/or PAM microarray was cloned and sequenced (Table S2). Most of the retrieved sequences corresponded to Betaproteobacteria (*Acidovorax*, *Comamonas*, and *Dechloromonas* genus), Gammaproteobacteria (Pseudomonadaceae family), and Acidobacteria divisions, which dominated the upper (90–120 mbs) and the lower (140–164 mbs) part of the anaerobic zone. In the intermediate, low sulfate region of this borehole (120–139 mbs), the sequences corresponded to Firmicutes and Actinobacteria phyla. Additionally, several sequences associated with halophilic Euryarchaea were found, in agreement with the positive hybridization with Euryarchaea-specific probes on PAM microarray.

Metabolically active bacterial clusters

To demonstrate the actual presence of microbial cells in the IPB microniches, several core samples from BH8 were stained with DAPI and hybridized with EUB338, a eubacterial-specific probe to 16S rRNA. DAPI-stained samples revealed the presence of clusters of microbial morphologies bound to micro-holes in the rock in several deep core samples (Fig. 6A). Some of these clusters were also efficiently stained by FISH and CARD-FISH, indicating that the bacterial cells were indeed alive and metabolically active (Fig. 6B). In addition, scanning electron microscopy (SEM) analysis of some core samples showed the presence of carbonaceous networks attached to pyrite minerals that can be originated from microbial exopolymers (not shown). Unfortunately, while microbial clusters were indeed detected by microscopy in this study, their number was too low and their distribution too sparse to allow for a reliable quantification of the biomass in the IPB subsurface.

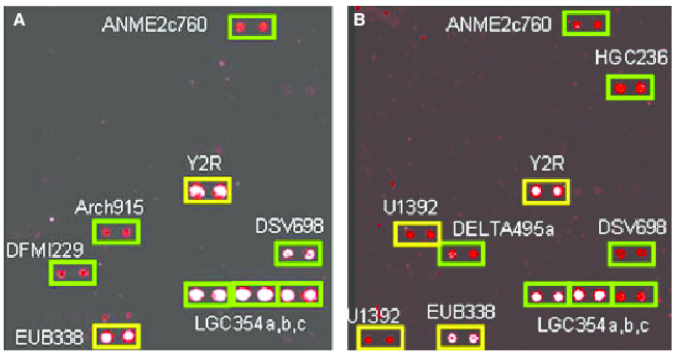


Fig. 5 Exploring the prokaryotic diversity of the IPB with an oligonucleotide microarray. Environmental DNA was extracted from different core samples, the 16S rRNA gene was amplified by PCR, fluorescently labeled and hybridized with PAM microarray (Materials and methods). Scanned fluorescent images obtained after hybridization of the amplified DNA from core samples 120.85 (A) and 126.85 mbs (B). Probes showing positive results (red to white spots) are indicated: EUB338 and Y2R, universal for Eubacteria; DFMI229, genus *Desulfotomaculum* (Firmicutes, Clostridiales); DSV698, sulfate reducer Deltaproteobacteria, Desulfovibrionaceae, most *Desulfovibrio* species and *Desulfomonas*; LGC354a, b, c, Low-GC Gram-positive Firmicutes *Lactobacillus* group (a), *Bacillus* group (b), and *Streptococcus* group (c); Arch915, Archaea; ANME2c760, Methanotrophic, Euryarchaeota; DELTA495a, Deltaproteobacteria and Gemmatimonadetes; HGC236, High-GC content Gram-positive Actinobacteria (Supporting Table S1 and Supporting Data S1).

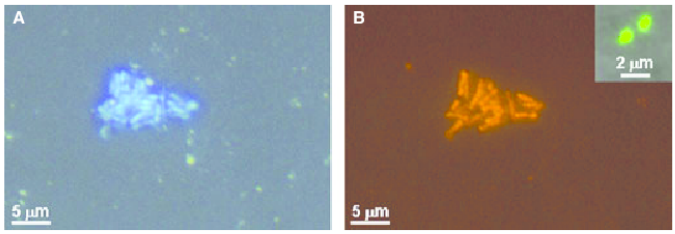


Fig. 6 Optical microscopy analysis of cores from the Iberian Pyrite Belt subsurface. (A), (B) Epifluorescence micrographs showing a eubacterial cell cluster at 154.7 mbs stained with DAPI (A), and the same field stained by FISH with the Eubacteria-specific probe EUB338 (B), and by CARD-FISH (insert in B) with the same probe.

DISCUSSION

A multitechnique approach to explore the geomicrobiology of a deep pyrite body

Although little is known about the small-scale distribution of micro-organisms in solid samples, some authors have stressed the difficulty of uniformly characterizing microbial communities in deep subsurface rocky samples (Haldeman *et al.*, 1993). Others have shown that in lacustrine sediments, microbial activities, such as sulfate reduction, are carried out in submillimeter microniches (Widerlund and Davidson, 2007). Such heterogeneity complicates the microbiological study of these types of samples. Also, the low concentration of microbes in the deep subsurface forces the sampling of a relatively high amount of material, especially in deep rocky samples (as in this work; Fig. 1), where the microbial distribution is dispersed throughout the mineral matrix (Sahl *et al.*, 2007; Von der Weid *et al.*, 2008). To overcome such a drawback, as well as the inherent bias of each analytical method, we retrieved many small (1–2 g) powder samples at short distances and analyzed them with multiple non-culture-dependent techniques (Materials and methods).

First, a biomarker and life detection experiment was performed by using antibody microarray-based biosensor technology. The advantages of LDChip are: (i) that the assay can be performed in the field, just after sampling, which reduces the chances of sample alteration and/or contamination; (ii) very little sample processing is required, just homogenization by a hand-held sonicator, filtering, and incubation; and (iii) up to hundreds of targets (microbes, proteins, and other biomolecules) can be assayed simultaneously. Its most critical limitations concern the cross-reaction events in a multiplex assay with a multianalyte-containing sample. We have addressed this issue by developing antibody graphs and a deconvolution method to distinguish the specific recognition events between antibodies and their cognate antigens from the cross-reaction with other related antigens (Rivas *et al.*, 2011). We concluded that although cross-reactions take place, they correspond to similar antigens or they come from phylogenetically related microbes. Another critical aspect of multiplex sandwich-type multiplex immunoassay is the inherent

background due to the high concentration of the revealing antibody mixture, in this case 200 antibodies. However, we demonstrated that the system still keeps a reliable dynamic range (Rivas *et al.*, 2008). In the present work, LDChip200 mainly detected biological polymers from the Gram-positive bacteria, mostly lipoteichoic acids and peptidoglycan. One of the positive reactions corresponded to an antiferritin antibody, indicating the presence of bacterioferritins or the highly related DPS proteins. These proteins form part of the universal 'Ferritin-like superfamily' that plays a critical role as cellular repository of the excess of iron (Andrews, 2010). They might be induced in the IPB subsurface as a consequence of the high iron concentration that micro-organisms have to deal with. Indeed, ferritins might be playing a detoxifying role by oxidizing Fe^{2+} (highly toxic) to Fe^{3+} and storing it inside the pockets these proteins can form (Andrews, 1998; Smith, 2004). Antibodies to crude environmental extracts from cores at different depths (ID7-12S2) also showed positive reactions with a sandwich-type immunoassay, indicating the presence of complex immunogenic polymers. Whether these polymers came from Gram-positive bacteria has yet to be determined. However, the oligonucleotide microarray clearly revealed the presence of DNA from the Gram-positive Firmicutes phylum and from the genus *Desulfotomaculum*, a group of obligate anaerobe sulfate-reducing bacteria (SRB). Interestingly, the LDChip200 detected *Acidithiobacillus caldus* or a related strain in a sample retrieved from 156.4 mbs (not shown), and its presence was corroborated by the positive DNA hybridization with the specific oligonucleotide probe THC642 (Fig. 7). We confirmed once again the usefulness of the LDChip immunosensor for the detection of biomarkers in the subsurface, and its tremendous potential for the search for life on Mars as the key component of Signs Of Life Detector (SOLID) instrument (Parro *et al.*, 2011). The SOLID is an instrument designed and constructed for the *in situ* search for molecular biomarkers on Mars (Parro *et al.*, 2005; 2008; Parro *et al.*, 2011). The instrument can extract the organic matter present in up to 0.5 g of dust, martian regolith, or ground rock into a liquid solution or suspension by means of ultrasonication. After a filtering process, the liquid filtrate floods a flow chamber containing the multiplex immunosensor LDChip. After incubation and washing steps, a mixture of fluorescently labeled antibodies, a laser beam to excite the fluorochrome, and a charge-coupled device (CCD) camera is used to reveal the positive antigen–antibody reactions. The results shown here indicate that the SOLID's life detection system works well with rocky deep subsurface samples from a martian analogue.

The LDChip200 showed a widespread signal from Gram-positive antigens, in agreement with the positive hybridization obtained with oligonucleotide probes for low- and high-GC-content Gram-positive bacteria with the PAM microarray. The PAM contained more probes than LDChip200, whose probe composition was highly biased for detecting micro-organisms from surface environments and lacked antibodies from many groups of prokaryotes. Although the use of multiple displacement amplification (MDA) for the PAM microarray hybridization resulted in lower fluorescent signal and higher background, it corroborated the results obtained with 16S rRNA gene PCRs with DNA extracts not subjected to MDA. Furthermore, it revealed the presence of additional microbial groups. This was the case of the probe to *Gallionella*, a genus from the Betaproteobacteria division that includes iron-oxidizing chemolithotrophs operating at low oxygen concentrations. It also revealed the presence of some members of methanogenic Archaea from the Euryarchaeota phylum, particularly from the Methanosarcina family (MSSH859) and Methanococcales order (MC1109), and members of the Deltaproteobacteria *Desulfosarcina/Desulfococcus* genera.

Some of the obtained 16S rRNA gene sequences were in agreement with the results from the immunological and hybridization techniques. This technique also revealed the presence of denitrifying Betaproteobacteria, which were not detected with the previous techniques. The reason for such a bias in the detection of Betaproteobacteria by 16S rRNA gene sequencing compared with the other two techniques may reside in the absence of antibodies in LDChip200 and DNA probes in the oligonucleotide microarray for this phylum (only a generic probe whose target is in 23S rRNA, BET42a) and/or to the natural bias produced in the PCR amplification in natural samples with very little DNA amounts. Furthermore, MDA is known to produce biases in community composition (Yilmaz *et al.*, 2010).

The detection by SEM and FISH of clusters of microbial cells bound to minerals supported the results obtained with the techniques shown above. Although it is impossible to completely exclude the possibility of contamination of the samples with drilling fluids, the absence of the bromide tracer in the analyzed samples, as well as microscopy showing the presence of mineral-bound microbial clusters, indicated that the molecular results indeed corresponded to indigenous micro-organisms of the IPB.

The fact that the results obtained with the different techniques did not match exactly in every single 0.5–1 g of core sample can be a consequence of the heterogeneity of rocky samples, as well as the intrinsic biases of each of the current molecular ecology techniques used (Polz & Cavanaugh, 1998; Yilmaz *et al.*, 2010). In spite of this limitation, the geochemical and biodiversity analyses allowed us to cluster the samples in differentiated geomicrobiological units (Fig. 8; see discussion below).

In conclusion, the microbiological analysis of low-biomass-containing environmental samples, such as those from deep drilling projects, requires the application of several complementary molecular techniques to obtain a more

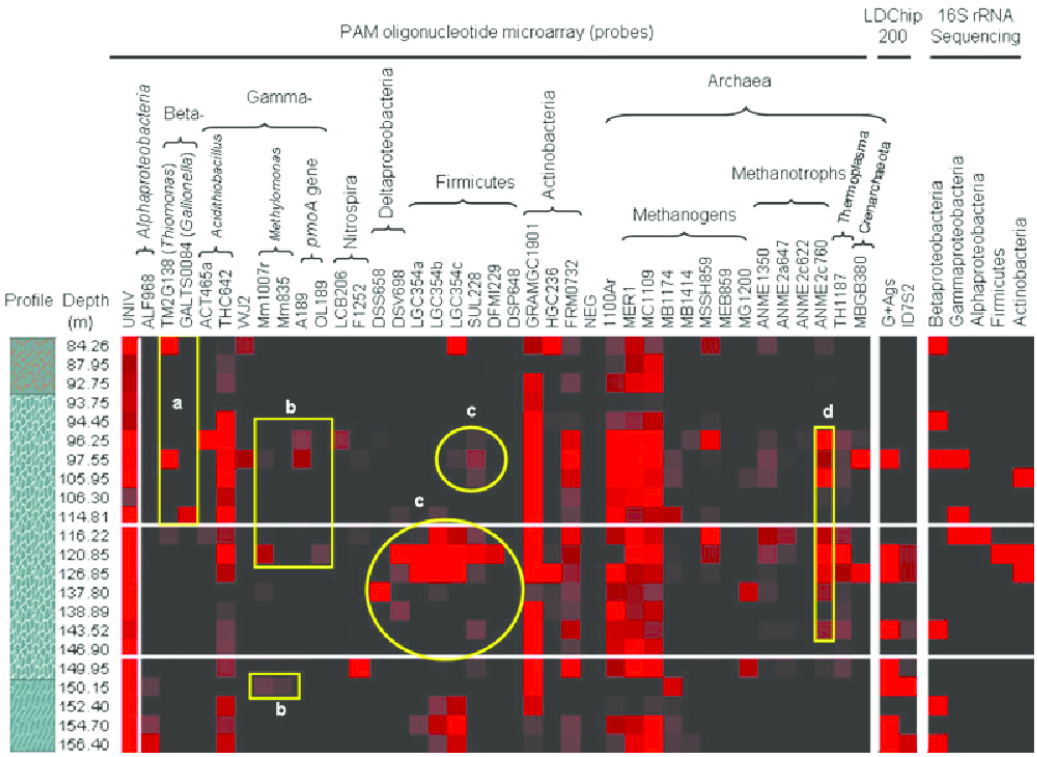


Fig. 7 Comparison between the results obtained with three different molecular techniques: Map of relative fluorescent intensities obtained after hybridization of the PCR-amplified 16S rRNA gene and total MDA-amplified metagenome with PAM oligonucleotide microarray; the fluorescent intensity of the most relevant positive biomarkers detected with the immunosensor LDChip200; and a positive (red) or negative detection (black) map of the retrieved 16S rRNA gene sequences. Each horizontal row of red and black squares represents the result for each probe obtained for a sample in the same chip. The geological profile and the depths at which the core samples were collected for analysis are indicated (left). The specificities of the PAM probes are detailed in the supplementary materials (Table S1). The LDChip200 probes listed in this figure are the following: GP_Ag, Gram-positive bacteria antigens (mostly lipoteichoic acids); ID7S2, immunoreactive EPS material from some of the cores. Based on the prokaryotic diversity results, three geomicrobiological levels can be distinguished (white horizontal lines), which are further discussed in the text. Letters from 'a' to 'h' indicate clusters of probes with differentiated patterns whose distribution is further discussed in the text.

precise picture of the microbial population and the operating metabolisms. The multitechnique approach we have shown here, combining *in situ* immunosensing, multiple displacement metagenome amplification, PCR, oligonucleotide microarray hybridization, and sequencing, is particularly useful to complement the biases that PCR can produce in metagenomic studies, where, very often, the amount and the quality of the template is far from the ideal.

Potential microbial metabolisms in the deep IPB subsurface

Deep subsurface ecosystems are thought to be energy-limited, with low cell numbers subsisting at very low metabolic rates (Hoehler & Jørgensen, 2013). In subseafloor sediments, the readily degradable organic substrates disappear quickly with depth, leaving only recalcitrant compounds to support heterotrophic growth (Parkes *et al.*, 2000). Altogether, the geochemical and the immunological results indicated the presence of high-molecular-weight and complex organic matter that can originate from extant microbial cells (protein, polysaccharides) or from extinct biomass (aromatic compounds). Small organic acids such as acetate or formate could be used as efficient carbon and energy sources, while rock-water interactions in deep bedrock fracture environments could provide the hydrogen necessary to support litho-autotrophic metabolisms (Stevens & McKinley, 1995; Freund *et al.*, 2002).

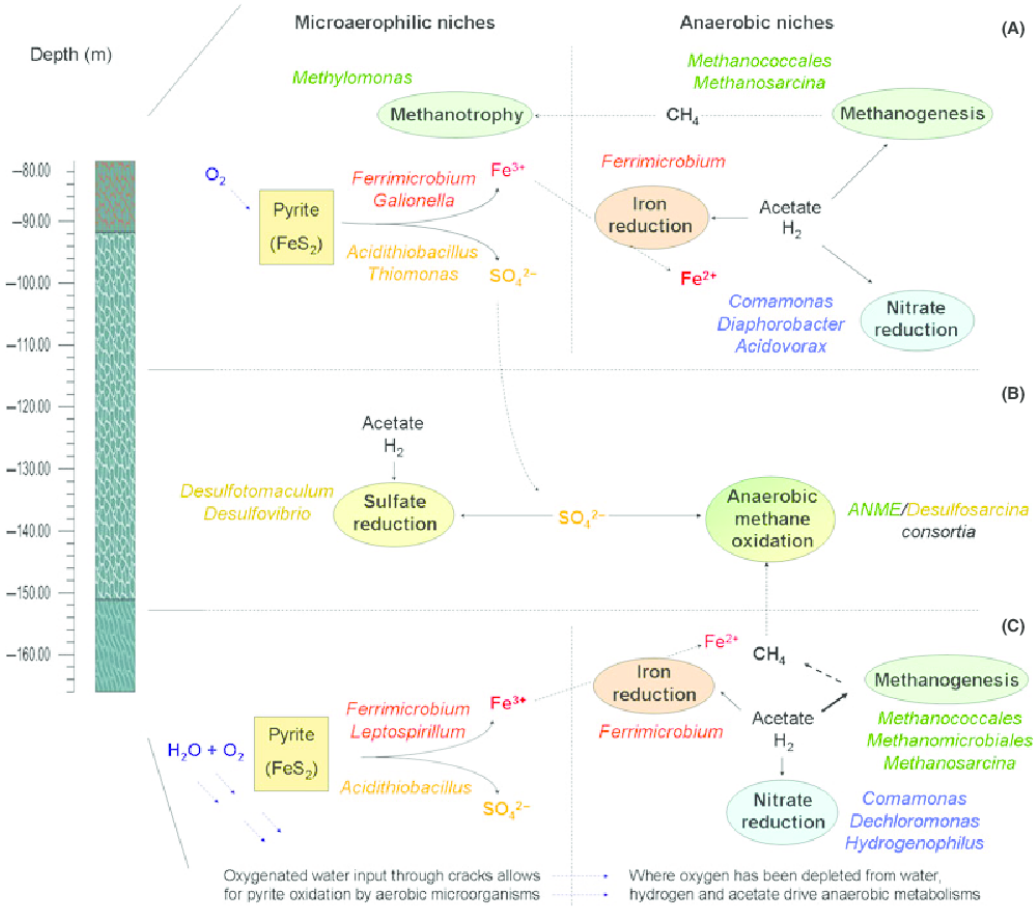


Fig. 8 Potential metabolisms operating in the IPB subsurface. Based on the geochemical, immunological, DNA hybridization, and sequencing results, we propose a three-compartment model for the deep IPB subsurface ecosystem: (A) Microaerophilic zone (84–116 mbs); (B) Anaerobic zone (115–149 mbs); and (C) Fractured zone (150–164 mbs). See Discussion for details.

Water availability is also a key factor in subsurface environments, as it will strongly affect microbial metabolic rates and diffusion of molecular species. While most of the core samples retrieved in this work were generally compact and non-porous, the presence of cracks in some of the deepest cores (below 150 m; Fig. 2) would result in an increase in the water influx, as well as in the space available for microbial colonization. Interestingly, a diverse prokaryotic community was detected in that zone (Fig. 8C), along with alterations in the geochemical profiles, such as an increase in sulfate and methane concentrations, and a decrease in ferrous iron (Fig. 2), which could be products of an increased microbial activity. The presence of iron oxides and iron-oxidizing micro-organisms (*Leptospirillum* spp. and *Acidithiobacillus* spp., Fig. 7) in this zone also suggests that the water entering the system carries dissolved oxygen. Such oxygen would be readily consumed by these pyrite-oxidizing microbes, rendering an anoxic fluid charged with sulfate and ferric iron. This would provide an input of oxidants to the anaerobic part of the community, which contains sulfate- and iron-reducing micro-organisms. In addition, methanogens were also detected (Fig. 7, probes MER1, MG1200, MC1109, MSSH859), together with an increase in methane concentration (Fig. 2). The co-occurrence of high sulfate and methane levels in this zone is interesting, as sulfate is known to inhibit methanogenesis (Winfrey & Zeikus, 1977; Abram & Nedwell, 1978) via substrate competition

from SRB. In spite of that, there was no detection of SRB below 150 mbs, while methanogens were found in almost every sample. The reason for the predominance of methanogenesis over sulfate reduction is unclear, but it has been described that SRB can be inhibited by high heavy metal concentrations and precipitation of metal sulfides (Utgikar *et al.*, 2002).

Ferric iron is another inhibitor of methanogenesis (Lovley & Phillips, 1987; Bodegom *et al.*, 2004; Zhang *et al.*, 2009). Nevertheless, Sanz *et al.* (2011) reported methanogenic activity in microcosms from ferric iron-rich sediments of the Tinto River, as well as in *in situ* measurements. They concluded that methanogenesis was being carried out in reduced microniches generated by iron-respiring bacteria. Our results support this hypothesis in the sense that the distribution of methanogens and ferric iron reducers (e.g. *Ferrimicrobium* spp.) followed a similar pattern throughout the whole borehole (Fig. 7, probes FRM0732, MER1, MC1109).

The fact that methane was only detected below 140 mbs in spite of methanogens being ubiquitous in the borehole can be explained by the presence of methane oxidizers in all but the deepest samples (Fig. 7b,d). Above 149 mbs, there seems to be a close association between anaerobic methane oxidizers (ANMEs) and SRB (Fig. 7c,d), which could mean that methane oxidation coupled to sulfate reduction (Orphan *et al.*, 2001; Parkes *et al.*, 2007; Knittel & Boetius, 2009; Pedersen, 2013) would be the main methane sink in the borehole (Fig 8B).

The presence of 16S rRNA sequences from denitrifying bacteria (Fig. 7, Betaproteobacteria) is more difficult to link to the geochemical data, as the nitrate and nitrite distributions do not seem to follow a clear pattern. Nitrite peaks observed at several depths could be interpreted as the product of nitrate reduction, although their distribution is not entirely consistent with the detection of denitrifiers in the borehole.

Because the water table was located at 90 m below the surface, we assume a smooth transition from an aerobic/microaerobic level to an anaerobic one (below 115 m deep). Detection of aerobic and microaerophillic pyrite oxidizers such as *Galionella* spp. and *Thiomonas* spp. above 115 mbs (Fig. 7a) does in fact suggest that some aerobic microniches exist in that zone (Fig. 8A). In this part of the drill, microorganisms such as *Gallionella* spp might be oxidizing metals (Fe^{2+} and Mn^{2+}), and others such as *Thiomonas* spp and *Acidithiobacillus* spp could oxidize Fe^{2+} and reduced sulfur compounds (e.g. thiosulfate, $S_2O_3^{2-}$). Moreover, in those microaerobic niches, the metanotrophic bacteria *Methylomonas* spp. (Fig. 7b) could account for the aerobic oxidation of methane while anaerobic metanotrophs (Fig. 7d) would do the same in the anaerobic niches.

The integration of geochemical, immunological, DNA hybridization, and sequencing results, allowed us to propose a three-compartment model for the deep IPB subsurface ecosystem (Fig. 8); (i) a microaerophilic zone (84–116 mbs), in which methane, iron, and sulfur (including pyrite, FeS_2) are oxidized in aerobic microniches, while denitrification and methanogenesis associated with ferric iron reduction occurs in the anaerobic ones; (ii) an anaerobic zone (115–149 mbs), in which anaerobic methane oxidation might be coupled to sulfate reduction; and (iii) a fractured zone (150–164 mbs), in which an increased water and oxygen availability would allow for higher metabolic rates. Pyrite oxidation would consume dissolved oxygen, generating anaerobic microniches in which methanogenesis would be the main process. Thus, microbial activities would be responsible for the disappearance of reduced iron and the generation of sulfate and methane at those depths (Fig. 2).

In conclusion, we showed evidence for a rich prokaryotic diversity operating in the deep subsurface of the massive sulfide ores of the IPB. The LDChip200 immunosensor identified microbial polymers from Gram-positive bacteria as well as some proteins from the ferritin family at different depths. This result was confirmed with the PAM oligonucleotide microarray, which detected Gram-positive bacteria from the Firmicutes and Actinobacteria phyla. Additionally, PAM detected sulfate-reducing bacteria, methanogenic Archaea, anaerobic methane oxidizers, as well as bacteria involved in the iron/sulfur cycle. These results allowed us to infer a model of the deep subsurface habitat of the IPB (Fig. 8). However, many questions are still unanswered: the origin of acetate and formate remains to be clarified, the presence of acetogenic and hydrogenotrophic bacteria must be confirmed, and other metabolic traits not yet identified should be searched for. A deeper molecular and phylogenetic analysis of the microbial diversity along the borehole will contribute to clarify some of these questions. The European Research Council has granted the IPB Subsurface Life (IPBSL) project, an Advanced Grant project, now underway to perform deeper drillings and address these and other questions concerning the geomicrobiology of this unique massive pyrite site.

ACKNOWLEDGMENTS

We thank all MARTE team members. The MARTE project (Mars Astrobiology Research and Technology Experiment) was a NASA-CAB joint project funded by the NASA ASTEP program project No. NRA-02-OSS-01 and by the CAB internal budget. The present work was supported by the *Subdirección General de Proyectos de Investigación*, of the Spanish *Ministerio de Economía y Competitividad* (MINECO) Grants No. ESP2004-05008, ESP2006-08128, AYA2008-04013, and the European Research Council (ERC) Advanced Grant No. 250350. F. Puente-Sánchez has a JAE-pre fellowship from the Spanish *Consejo Superior*

de Investigaciones Científicas (CSIC). We also thank Dr. Enoma Omoregie for revising the manuscript and providing useful comments. There are no conflicts of interests.

REFERENCES

Abram JW, Nedwell DB (1978) Inhibition of methanogenesis by sulphate reducing bacteria competing for transferred hydrogen. *Archives of Microbiology* **117**, 89–92.

Amann RI (1995) In situ identification of microorganisms by whole cell hybridization with rRNA-targeted nucleic acid probes. In *Molecular Microbial Ecology Manual* (eds Akkermans ADL, van Elsas JD, de Bruijn FJ). Kluwer Academic Publishers, Dordrecht, pp. 1–15.

Amils R, González-Toril E, Fernández-Remolar DC, Gómez F, Aguilera A, Rodríguez N, Malki M, García-Moyano A, Fairén AG, de la Fuente V, Sanz JL (2007) Extreme environments as Mars terrestrial analogs: the Rio Tinto case. *Planetary and Space Science* **55**, 370–381.

Andrews SC (1998) Iron storage in bacteria. *Advances in Microbial Physiology* **40**, 281–351.

Andrews SC (2010) The Ferritin-like superfamily: evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica et Biophysica Acta* **1800**, 691–705.

Blanco Y, Prieto-Ballesteros O, Gómez MJ, Moreno-Paz M, García-Villadangos M, Rodríguez-Manfredi JA, Cruz-Gil P, Sánchez-Román M, Rivas LA, Parro V (2012) Prokaryotic communities and operating metabolisms in the surface and the permafrost of Deception Island (Antarctica). *Environmental Microbiology* **14**, 2495–2510.

Bodegom PM, Scholten JC, Stams AJ (2004) Direct inhibition of methanogenesis by ferric iron. *FEMS Microbiology Ecology* **49**, 261–268.

Daims H, Bruhl A, Amann R, Schleifer K-H, Wagner M (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and Applied Microbiology* **22**, 434–444.

Fernández-Remolar DC, Knoll AH (2008) Fossilization potential of iron-bearing minerals in acidic environments of Rio Tinto, Spain: implications for Mars exploration. *Icarus* **194**, 72–85.

Fernández-Remolar DC, Morris RV, Gruener JE, Amils R, Knoll AH (2005) The Río Tinto Basin, Spain: mineralogy, sedimentary geobiology, and implications for interpretation of outcrop rocks at Meridiani Planum, Mars. *Earth and Planetary Science Letters* **240**, 149–167.

Fernández-Remolar DC, Gómez F, Prieto-Ballesteros O, Schelble RT, Rodríguez N, Amils R (2008a) Some ecological mechanisms to generate habitability in planetary subsurface areas by chemolithotrophic communities: the río tinto subsurface ecosystem as a model system. *Astrobiology* **8**, 157–173.

Fernández-Remolar DC, Prieto-Ballesteros O, Rodríguez N, Gómez F, Amils R, Gómez-Elvira J, Stoker CR (2008b) Underground habitats in the Río Tinto basin: a model for subsurface life habitats on Mars. *Astrobiology* **8**, 1023–1047.

Fernández-Remolar DC, Preston LJ, Sánchez-Román M, Izawa MRM, Huang L, Southam G, Banerjee NR, Osinski GR, Flemming R, Gómez-Ortíz D, Prieto Ballesteros O, Rodríguez N, Amils R, Darby Dyar M (2012) Carbonate precipitation under bulk acidic conditions as a potential biosignature for searching life on Mars. *Earth and Planetary Science Letters* **351**, 13–26.

Freund F, Dickinson JT, Cash M (2002) Hydrogen in Rocks: an energy source for deep microbial communities. *Astrobiology* **2**, 83–92.

Garrido P, González-Toril E, García-Moyano A, Moreno-Paz M, Amils R, Parro V (2008) An oligonucleotide prokaryotic acidophile microarray: its validation and its use to monitor seasonal variations in extreme acidic environments with total environmental RNA. *Environmental Microbiology* **10**, 836–850.

Haldeman DL, Amy PS, Ringelberg D, White DC (1993) Characterization of the microbiology within a 21 m^3 section of rock from the deep subsurface. *Microbial Ecology* **26**, 145–159.

Hoehler TM, Jørgensen BB (2013) Microbial life under extreme energy limitation. *Nature Reviews Microbiology* **11**, 83–94.

Knittel K, Boetius A (2009) Anaerobic oxidation of methane: progress with an unknown process. *Annual Review of Microbiology* **63**, 311–334.

Leistel JM, Marcoux E, Thieblemont D, Quesada C, Sánchez A, Almodóvar GR, Pascual E, Sáez R (1997) The volcanic-hosted massive sulphide deposits of the Iberian Pyrite Belt. *Mineralium Deposita* **33**, 2–30.

Lovley DR, Phillips EJ (1987) Competitive mechanisms for inhibition of sulfate reduction and methane production in the zone of ferric iron reduction in sediments. *Applied and Environment Microbiology* **53**, 2636–2641.

Orphan VJ, Hinrichs KU, Ussler W 3rd, Paull CK, Taylor LT, Sylva SP, Hayes JM, Delong EF (2001) Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Applied and Environment Microbiology* **67**, 1922–1934.

Parkes RJ, Cragg BA, Wellsbury P (2000) Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeology Journal* **8**, 11–28.

Parkes RJ, Cragg BA, Banning N, Brock F, Webster G, Fry JC, Hornibrook E, Pancost RD, Kelly S, Knab N, Jørgensen BB, Rinna J, Weightman AJ (2007) Biogeochemistry and biodiversity of methane cycling in subsurface marine sediments (Skagerrak, Denmark). *Environmental Microbiology* **9**, 1146–1161.

Parro V, Rodríguez-Manfredi JA, Briones C, Compostizo C, Herrero PL, Vez E, Sebastián E, Moreno-Paz M, García-Villadangos M, Fernández-Calvo P, González-Toril E, Pérez-Mercader J, Fernández-Remolar D, Gómez-Elvira J (2005) Instrument development to search for biomarkers on mars: terrestrial acidophile, iron-powered chemolithoautotrophic communities as model systems. *Planet and Space Science* **53**, 729–737.

Parro V, Fernández-Calvo P, Rodríguez-Manfredi JA, Moreno-Paz M, Rivas LA, García-Villadangos M, Bonaccorsi R, González-Pastor JE, Prieto-Ballesteros O, Schuerger AC, Davidson M, Gómez-Elvira J, Stoker CR (2008) SOLID2: an antibody array-based life-detector instrument in a mars drilling simulation experiment (MARTE). *Astrobiology* **8**, 987–999.

Parro V (2010) Antibody microarrays for environmental monitoring. In *Handbook of Hydrocarbon and Lipid Microbiology* (ed Timmis KN). Springer-Verlag, Berlin, pp. 2699–2710.

Parro V, De Diego-Castilla G, Moreno-Paz M, Blanco Y, Cruz-Gil P, Rodríguez-Manfredi JA, Fernández-Remolar D, Gómez F, Gómez MJ, Rivas LA, Demergasso C, Echeverría A, Urtuvia VN, Ruiz-Bermejo M, García-Villadangos M, Postigo M, Sánchez-Román M, Chong-Díaz G, Gómez-Elvira J (2011) A microbial oasis in the hypersaline Atacama subsurface discovered by a life detector chip: implications for the search for life on Mars. *Astrobiology* **11**, 969–996.

Pedersen K (2013) Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H(2). *ISME Journal* **7**, 839–849.
Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environment Microbiology* **68**, 3094–3101.
Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multi-template PCR. *Applied and Environment Microbiology* **64**, 3724–3730.
Rivas LA, García-Villadangos M, Moreno-Paz M, Cruz-Gil P, Gómez-Elvira P, Parro V (2008) A 200-antibody microarray biochip for environmental monitoring: searching for universal microbial biomarkers through immunoprofiling. *Analytical Chemistry* **80**, 7970–7979.
Rivas LA, Aguirre J, Blanco Y, González-Toril E, Parro V (2011) Graph-based deconvolution analysis of multiplex sandwich microarray immunoassays: applications for environmental monitoring. *Environmental Microbiology* **13**, 1421–1432.
Sahl JW, Schmidt R, Swanner ED, Mandernack KW, Templeton AS, Kieft TL, Smith RL, Sanford WE, Callaghan RL, Mitton JB, Spear JR (2007) Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Applied and Environment Microbiology* **74**, 143–152.
Sanz JL, Rodríguez N, Díaz EE, Amils R (2011) Methanogenesis in the sediments of Rio Tinto, an extreme acidic river. *Environmental Microbiology* **13**, 2336–2341.
Smith JL (2004) The physiological role of ferritin-like compounds in bacteria. *Critical Reviews in Microbiology* **30**, 173–185.
Stahl DA, Amann R (1991) Development and application of nucleic acid probes. In *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt E, Goodfellow M). Wiley, Chichiester, pp. 205–248.
Stevens TO, McKinley JP (1995) Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* **270**, 450–455.
Stoker CR, Cannon HN, Dunagan SE, Lemke LG, Glass BJ, Miller D, Gomez-Elvira J, Davis K, Zavaleta J, Winterholler A, Roman M, Rodriguez-Manfredi JA, Bonaccorsi R, Bell MS, Brown A, Battler M, Chen B, Cooper G, Davidson M, Fernández-Remolar D, González-Pastor E, Heldmann JL, Martínez-Frías J, Parro V, Prieto-Ballesteros O, Sutter B, Schuerger AC, Schutt J, Rull F (2008) The 2005 MARTE robotic drilling experiment in Río Tinto, Spain: objectives, approach, and results of a simulated mission to search for life in the martian subsurface. *Astrobiology* **8**, 921–945.
Tuross N, Stathoplos L (1993) Ancient proteins in fossil bones. *Methods in Enzymology* **224**, 121–129.
Utgikar VP, Harmon SM, Chaudhary N, Tabak HH, Govind R, Haines JR (2002) Inhibition of sulfate-reducing bacteria by metal sulfide formation in bioremediation of acid mine drainage. *Environmental Toxicology* **17**, 40–48.
Von der Weid I, Korenblum E, Jurelevicius D, Rosado AS, Dino R, Sebastian GV, Seldin L (2008) Molecular diversity of bacterial communities from subseafloor rock samples in a deep-water production basin in Brazil. *Journal of Microbiology and Biotechnology* **18**, 5–14.
Widerlund A, Davison W (2007) Size and density distribution of sulfide-producing microniches in lake sediments. *Environmental Science and Technology* **41**, 8044–8049.
Winfrey MR, Zeikus JG (1977) Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments. *Applied and Environment Microbiology* **33**, 275–281.
Yilmaz S, Allgaier M, Hugenholtz P (2010) Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* **7**, 943–944.
Zhang L, Keller J, Yuan Z (2009) Inhibition of sulfate-reducing and methanogenic activities of anaerobic sewer biofilms by ferric iron dosing. *Water Research* **43**, 4123–4132.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Probes used for developing the Prokaryotic Acidophile Microarray (PAM) (Garrido *et al.*, 2008).

Table S2 Highest sequence identity scores of obtained 16S rRNA gene sequences to sequences from isolates deposited in the GeneBank database.

Table S3 Nomenclature of the main compounds detected by QTOF spectrometry.

Data S1 Microarray data: Fluorescence intensities obtained with PAM microarray after hybridizattion with DNA amplified from different core samples.

Río Tinto: a geochemical and mineralogical terrestrial analogue of mars

Life **2014**, *4*, 511-534; doi:10.3390/life4030511

OPEN ACCESS

life

ISSN 2075-1729

www.mdpi.com/journal/life

Review

Río Tinto: A Geochemical and Mineralogical Terrestrial Analogue of Mars

Ricardo Amils [1,2,*], David Fernández-Remolar [2] and the IPBSL Team [†]

[1] Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

[2] Centro de Astrobiología (CSIC-INTA), km 4 carrtera Ajalvir, 28850 Torrejón de Ardoz, Spain; E-Mail: kernnunnos@gmail.com

[†] the IPBSL Team: Victor Parro, José Antonio Rodríguez-Manfredi, Monike Oggerin, Mónica Sánchez-Román, Francisco J. López, José Pablo Fernández-Rodríguez, Fernando Puente-Sánchez, Carlos Briones, Olga Prieto-Ballesteros, Fernando Tornos, Felipe Gómez, Miriam García-Villadangos, Nuria Rodríguez, Enoma Omoregie (*Centro de Astrobiología, carretera de Ajalvir km 4, 28850 Torrejón de Ardoz, Spain*); Kenneth Timmis, Alejandro Arce (*Institute of Microbiology, Technische Universitat Braunschweig, D38106 Braunschweig, Germany*); José Luis Sanz (*Departamento de Biología Molecular, Universidad Autónoma de Madrid, Cantoblanco, 28049, Madrid, Spain*); David Gómez-Ortiz (*ESCET-Area de Geología, Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain*).

* Author to whom correspondence should be addressed; E-Mail: ramils@cbm.csic.es; Tel.: +34-911-96-45-04; Fax: +34-911-96-45-44.

Received: 8 July 2014; in revised form: 22 August 2014 / Accepted: 28 August 2014 / Published: 15 September 2014

Abstract: The geomicrobiological characterization of the water column and sediments of Río Tinto (Huelva, Southwestern Spain) have proven the importance of the iron and the sulfur cycles, not only in generating the extreme conditions of the habitat (low pH, high concentration of toxic heavy metals), but also in maintaining the high level of microbial diversity detected in the basin. It has been proven that the extreme acidic conditions of Río Tinto basin are not the product of 5000 years of mining activity in the area, but the consequence of an active underground bioreactor that obtains its energy from the massive sulfidic minerals existing in the Iberian Pyrite Belt. Two drilling projects, MARTE (Mars Astrobiology Research and Technology Experiment) (2003–2006) and IPBSL (Iberian Pyrite Belt Subsurface Life Detection) (2011–2015), were developed and carried out to

provide evidence of subsurface microbial activity and the potential resources that support these activities. The reduced substrates and the oxidants that drive the system appear to come from the rock matrix. These resources need only groundwater to launch diverse microbial metabolisms. The similarities between the vast sulfate and iron oxide deposits on Mars and the main sulfide bioleaching products found in the Tinto basin have given Río Tinto the status of a geochemical and mineralogical Mars terrestrial analogue.

Keywords: acidophiles; Río Tinto; Iberian Pyrite Belt; metal sulfides; iron oxidation; iron cycle; sulfur cycle; iron minerals; jarosite; Mars

1. Introduction

The NASA Astrobiology roadmap [1] highlights the interest in extreme environments and the microorganisms that live in them in evaluating the possible existence of life beyond Earth. Acidophiles are of special interest, because the environments in which they thrive are the product of the chemolithotrophic metabolism of microorganisms that obtain energy from reduced mineral substrates and are not adaptations to geophysical constraints, as with most extremophiles. The meager requirements they have places them among the best candidates for a successful primitive energy conservation system.

The Viking mission, considered the first astrobiological mission devoted to the search for signs of life on Mars, concluded that life had little chance of developing there given the extreme conditions detected on its surface [2]. In the last forty years, important advances in microbiology, mainly in the characterization of extreme environments, have challenged this pessimistic point of view. Thanks to research on extremophiles, we now know that life is extremely robust and adapts rapidly to different conditions.

Natural acidic environments have two major origins. One associated with volcanic activities, the sulfur world, and the other linked to mining activities. Coal and metal mining operations expose sulfidic minerals to the action of aerobic chemolithotrophic microorganisms, facilitating their growth and generating acid mine drainage (AMD) or acid rock drainage (ARD), which are the cause of important environmental problems [3].

The mechanism by which microbes obtain energy by oxidizing sulfide mineral has been controversial for many years [4], but the demonstration that ferric iron present in the cell envelopes of leaching microorganisms is responsible for the electron transfer from insoluble mineral substrates to the microbial electron transport chain has clarified the situation [5]. The differences observed by using various sulfide minerals are determined by chemical oxidation mechanisms, which depend on the crystallographic structure of the mineral substrates [6].

The acidophilic strict chemolithotroph, *Acidithiobacillus ferrooxidans* (formerly *Thiobacillus ferrooxidans*), was first isolated from a coal mine in the 1940s [7]. Although *At. ferrooxidans* can obtain energy b oxidizing either reduced sulfur compounds or ferrous iron, bioenergetic considerations gave more prominence to the sulfide oxidation reaction [8,9]. The discovery that some strict chemolithotrophs, such as *Leptospirillum ferrooxidans*, thrive using ferrous iron as their only source of energy and that their role in bioleaching operations and the generation of AMDs is much more important than previously thought has completely changed this point of view [10]. Furthermore, it is

well established that iron can be oxidized not only aerobically, but anaerobically, coupled to anoxygenic photosynthesis, using ferrous iron as environmental reducing power or anaerobic respiration using nitrate as an electron acceptor (denitrification) [11,12].

The recent discovery of subsurface chemolithotrophic microorganisms participating in a radiation-free biosphere has opened an interesting perspective in astrobiology [13–16]. There is a growing list of alternative sources of lithotrophic substrates (Fe^{2+}, S^{2-}, S^0, As^{3+}, Mn^{2+}, *etc.*), which widens the range of metabolic versatility of this energy conservation system. Furthermore, sulfur- and iron-oxidizing microorganisms coupled to sulfur- and iron-reducers have a critical role in the maintenance of the sulfur and iron cycles, two fundamental biogeochemical cycles.

Acidic environments vary greatly in their physicochemical characteristics and microbial ecology. High temperatures may be the result of biological activity, facilitating colonization by thermotolerant and thermophilic acidophiles. Acidic ecosystems associated with mining activities are, at the geological scale, very young. However, some metal mines have a rather long history. Mines such as Río Tinto are known to have been in operation more than 5000 years ago [17].

2. Rio Tinto

Río Tinto (Figure 1) is an unusual ecosystem due to its acidity (mean pH 2.3), size (92 km long), high concentration of heavy metals (Fe, Cu, Zn, As, *etc.*) and unexpectedly high level of microbial diversity [18–20]. Río Tinto springs up in Peña de Hierro, in the core of the Iberian Pyrite Belt (IPB), and flows into the Atlantic Ocean at Huelva. The IPB is one of the largest sulfidic deposits in the world. Massive bodies of iron and copper sulfides, as well as minor quantities of lead and zinc sulfides constitute the main mineral ores. Its formation by hydrothermalism took place during the Hercynian orogenesis [21–23].

Figure 1. Río Tinto basin at Berrocal (J. Segura).

The basin of the river covers an area of 1700 km^2. The measured redox potentials in the Tinto basin range from +280 to +650 mV, and the oxygen content varies from saturation to complete anoxic conditions. The Tinto basin exhibits a Mediterranean climate. An important characteristic of the Tinto

ecosystem is its constant acidic pH, which is a direct consequence of the strong buffer capacity of ferric iron. Due to its size and easy access, Río Tinto is considered an excellent model for the study of the microbial ecology of extreme acidic environments.

The combined use of conventional and molecular microbial ecology methodologies has led to the identification of the most representative Tinto basin prokaryotic microorganisms [18–20,24–26]. Eighty percent of the water column diversity corresponds to microorganisms belonging to three bacterial genera: *Acidithiobacillus*, *Leptospirillum* and *Acidiphilium*, all members of the iron cycle [20]. *At. ferrooxidans* can oxidize ferrous iron aerobically and reduce ferric iron in anaerobic conditions [27,28]. All *Leptospirillum* isolated from Río Tinto are aerobic iron oxidizers. All *Acidiphilium* isolates can oxidize organic compounds using ferric iron as an electron acceptor (anaerobic respiration). Some *Acidiphilium* isolates can use ferric iron as an electron acceptor in the presence of oxygen [29,30]. Although other iron-oxidizers (*Ferrovum* spp., *Ferrimicrobium* spp., *Ferroplasma* spp. and *Thermoplasma acidophilum*) or iron-reducers (*Ferrimicrobium* spp., *Acidisphaera* spp., *Metallibacterium* spp. and *Acidobacterium* spp.) have been identified in the Tinto basin [20,25,26], their low numbers detected by *in situ* fluorescence hybridization suggest that they play a minor role in the operation of the iron cycle, at least in the water column.

Concerning the sulfur cycle, only *At. ferrooxidans* is found in significant numbers in the water column. This bacterium can oxidize both ferrous iron and reduced sulfur compounds. Reduced sulfur compounds can be oxidized aerobically and anaerobically. Sulfate reducing microorganisms that close the sulfur cycle have been detected in the sediments in various locations along the river [26,28,31–33].

Due to the biotechnological interests of the aerobic iron oxidizing microorganisms, interest in the characterization of the anoxic sediments from acidic environments has been overshadowed until recently, with few exceptions [34]. Recently, a comparative analysis of the sediments and the water column of different samples along the physicochemical gradient of Río Tinto have been performed [26]. This study showed a significantly higher level of biodiversity in the anaerobic sediments when compared to their water column counterparts. Nearly all of the microorganisms identified in this study were, in one way or another, related to the iron cycle. Most had been previously detected and/or isolated in AMD sites [3,20,25,26,35,36] or at biohydrometallurgical operations [37]. Nonetheless, some bacteria, such as members of Actinobacteria, Firmicutes, Acidobacteria, Planctomycetes and Chloroflexi, have been identified recently in the Tinto basin [26].

An in-depth analysis of two anoxic sediments from the Tinto basin has shown major phylogenetic differences among sample sites [31]. In one of the sediments, JL Dam, the most numerous group of bacteria corresponded to the phylum, Firmicutes (56.6%), followed by the phylum, Acidobacteria (27.3%), and the class, Deltaproteobacteria (11.6%). In the SN Dam, Proteobacteria was the most represented phylum (72.1%), followed by Actinobacteria (20.4%), while organisms of the Firmicutes and Acidobacteria phyla were present in low percentages. In the strict anoxic conditions detected in the lower part of the sediments of the Tinto basin, sulfate reduction is a recurrent microbial activity, a consequence of the high concentration of sulfates existing in the system.

It is usually assumed that the toxicity of high metal concentrations in acidic habitats limits eukaryotic growth and diversity. However, colorful biofilms cover large surfaces of the Tinto basin. In fact, it has been observed that eukaryotic microorganisms contribute over 60% of the Tinto basin biomass [18]. A significant number of eukaryotic species thriving in Río Tinto are photosynthetic.

Among them, chlorophytes related to different genera, such as *Chlamydomonas*, *Dunaliella* and *Chlorella*, are the most abundant eukaryotic microorganisms in the river [19,38–40]. Filamentous algae, represented by the genera, *Zygnemopsis* and *Klebsormidium*, have been also found. The most extreme part of the river is inhabited by a eukaryotic community dominated by two species related to the genera, *Dunaliella* and *Cyanidium*. Pennate diatoms are also present in the river, forming large brown biofilms [38,39]. Photoautotrophic flagellates of the genera, *Euglena*, *Bodo* and *Ochromonas*, are also widely distributed along the river. The dominant ciliate taxa belong to the order *Hypotrichida*. Amoebas are frequently found feeding on large diatoms, even in the most acidic part of the river. Heliozoa seem to be the characteristic top predators of the benthic food chain [19,38,39]. The only animal found in the river is a species of bdelloid rotifer related to the genus, *Rotifera* [19].

Among decomposers, fungi are the most diverse, and both unicellular and filamentous forms are present [18,41]. A recent characterization of Río Tinto basin samples from different stations along the river have rendered more than three hundred and fifty fungal isolates, which have been identified by ITS region sequence analysis. This analysis revealed Ascomycetes as the most abundant phylum, while Basidiomycetes and Zygomycetes accounted for less than 2% of the sequenced isolates. Of the Ascomycetes, 52% clustered within the Eurotiomycetes class, while 27% grouped with the Dothideomycetes and 17% with the Sordariomycetes. Concerning metal tolerance, Eurotiomycetes and Sordariomycetes isolates showed, in general, a high level of resistance to toxic heavy metals, much higher than the concentrations detected in the river, while members of the Dothideomycetes showed a level of resistance to concentrations similar to those detected in the water column.

However, not only unicellular eukaryotic systems develop in the extreme conditions of Río Tinto. Different plants can be found growing in the acidic soils of the Tinto basin [42–44]. The strategies used by these plants to overcome the physiological problems associated with the extreme conditions of the habitat are diverse. Some are resistant to the high concentration of heavy metals present in the soils in which they grow. Others specifically concentrate metals in different plant tissues. The analysis of the iron minerals found in the rhizomes and leaves of *Imperata cylindrica*, an iron hyperaccumulator perennial grass growing in the Tinto basin, showed significant concentrations of jarosite and iron hydroxides [42,45,46]. These results suggest that the management of heavy metals, in general, and iron, in particular, is much more complex in plants than what has been described to date. Furthermore, these results prove that multicellular complex systems can also develop in some extreme conditions, like those existing in Río Tinto.

3. Iron Bioformations in the Tinto Basin

Most of the Tinto basin biomass is located on the surface of the rocks in the riverbed. It is made up of dense biofilms, composed mainly of filamentous algae and fungi in which prokaryotic microorganisms are trapped. Significant iron mineral precipitation occurs on the negatively-charged surface of these biofilms, generating iron precipitates, which grow following the hydrological cycles and consolidate as iron-rich deposits elevated above the present river in the form of fluvial terraces (Figure 2) [47,48].

Seasonal evaporation of river water drives precipitation of hydronium jarosite and schwertmannite, while copiapite, coquimbite, gypsum and other sulfate minerals generate efflorescence brought to the

surface by capillary action [48]. During the wet season, hydrolysis of sulfate salts added to the effect of iron hydrolysis, facilitating the precipitation of amorphous iron oxyhydroxides.

The oldest terraces show increasing goethite crystallinity and its replacement by hematite over time. Organic matter does not preserve well in the Río Tinto sediments, but biosignatures imparted to sedimentary rocks as macroscopic textures of coated microbial streamers, surface blisters originating from biogenic gas and microfossils preserved in iron oxides can help to shape strategies for their detection in extant or future space exploration missions [48–50]. Interestingly, the specific biomineralization of hydronium jarosite by a filamentous fungus isolated from the Tinto basin, *Purpureocillium lilacinum*, in non-permissive ionic conditions has been recently described [51]. Furthermore, the presence of siderite ($FeCO_3$) in the modern sediments of the river [52], which has also been generated in cultures of the acidophilic heterotrophic iron reducer, *Acidiphilium* sp., in acidic conditions [53], strongly suggests that the presence of biological nucleation sites can modify the expected mineral precipitation schemes offered by the bulk physicochemical conditions in which microorganisms grow.

The recent detection of protein fragments and other organic molecules [54,55] in the ancient terraces of the Río Tinto basin evidences that its acidic and ferruginous environment promotes the preservation of molecules bearing information about the producing organism that inhabited this extreme environment over the last million years.

Figure 2. Alto de la Mesa old terrace.

Until recently, it was generally accepted that the extreme conditions found in Río Tinto were the direct result of the mining activities performed in the area during the last 5000 years [17,56]. New geological, geophysical and hydrogeological information supports the hypothesis that this is not the case. The location of the recharge area of the Peña de Hierro aquifer was recently determined northwest of the pit lake at a depth ranging from −100 to −400 m [57]. The groundwater moves southwards along the fracture network, and when it reaches the remnants of massive and/or stockwork sulfide bodies, located at −500 m in depth, the water interacts with the mineral substrate, facilitating the metabolism of chemolithotrophic microorganisms and generating acidic fluids [57,58].

Groundwater is eventually pumped along open strike-slip normal faults to reach the surface, where it sources the acidic springs that feed the headwaters of Río Tinto. The generation of acidic water occurs naturally through the oxidation of subsurface sulfidic bodies. Therefore, mining is not necessarily the cause of the characteristic low pH and high concentration of metals found in the river. This hypothesis is strongly supported by the sedimentary record of the ancient terrace deposits along the Tinto basin, which largely predate the oldest mining activity in the area [47,48]. The oldest terrace, containing finely laminated, as well as massive ironstones, has been dated 2.1 Ma [48]. Furthermore, the gossan deposits in the area, representing the remaining materials of the *in situ* oxidation of the massive and stockwork sulfide deposits, have been dated as older than 6 Ma [59].

Considering the geomicrobiological characteristics of the Tinto ecosystem, we postulate that the river is predominantly under the control of iron [9]. Iron is the main product of bioleaching of pyrite and iron bearing minerals, like chalcopyrite, both present in high concentrations in the IPB. The activity of iron oxidizing microorganisms is responsible for both the solubilization of sulfidic minerals and the corresponding high concentrations of iron, sulfate and protons detected in the water column of the river.

4. Iron World

Iron has diverse properties of ecological interest that make the Tinto ecosystem an interesting focus for astrobiological studies. Iron is not only a source of energy as a chemolithotrophic electron donor in its reduced form (Fe^{2+}), but can also be used as an electron acceptor for anaerobic respiration in its oxidized form (Fe^{3+}). As mentioned above, ferric iron is responsible for controlling the pH of the ecosystem. Although the reaction is reversible, dilution by tributaries is stronger than evaporation, so an important part of the iron remains precipitated along the course of the river, giving rise to iron bioformations. Accordingly, the concentration of soluble iron decreases gradually from the origin to the mouth of the river. Furthermore, soluble ferric iron readily absorbs harmful UV radiation, protecting the organisms growing in its waters [60,61].

This iron-controlled scenario seems reasonable for the chemolithotrophic prokaryotes thriving in the Tinto ecosystem. However, given the high level of eukaryotic diversity detected in the Tinto basin [18,19] and the fact that most of the primary production of the system derives from the activity of eukaryotic photosynthetic protists, what is the advantage, if any, for eukaryotes to develop in an extreme acidic environment with high concentrations of toxic heavy metals?

A possible answer to this question may be linked to the limited availability of iron in a neutral world. Although iron is an extremely important element for life [4,62], it is a limiting factor for growth at neutral pH [63,64]. Organisms have developed very specific elaborate mechanisms to sequester iron anywhere they can find it [65,66]. Why is this so, when iron is one of the most abundant elements on Earth [4]? In an oxidizing atmosphere at neutral pH, soluble ferrous iron is rapidly oxidized into insoluble compounds, which are incorporated into anaerobic sediments, where sulfate reducing microorganisms may further transform them into pyrite, an even less reactive iron mineral at neutral pH. The geological recycling of these sediments and the microbiology associated with the iron cycle are the only ways to reintroduce this critical element into the biosphere. The possible advantage for the eukaryotes thriving in the extreme conditions of Río Tinto is an unlimited iron supply provided by the chemolithotrophs growing in the rich iron sulfides of the IPB [9,67]. The availability of iron and other

heavy metals is considered so important for life that a model for an anoxic Proterozoic ocean deprived of iron and other heavy metals as a consequence of intense sulfate reducing activity has been proposed [68].

5. Subsurface Geomicrobiology of the Iberian Pyrite Belt

From the results discussed so far, it is clear that the main characteristics of the Tinto basin are not the product of mining contamination, but a consequence of the existence of an underground bioreactor, in which the massive sulfidic minerals of the Iberian Pyrite Belt are the main energy source, feeding the river with the products of the metabolic reactions occurring in the subsurface. The MARTE (Mars Astrobiology Research and Technology Experiment) (2003–2006) and IPBSL (Iberian Pyrite Belt Subsurface Life Detection) (2011–2015) drilling projects were designed to test this hypothesis by intersecting this underground reactor to provide evidence of subsurface microbial activities.

The main goal of the MARTE project (Mars Astrobiology Research and Technology Experiment), a collaborative effort between NASA and the Centro de Astrobiología, was the search for subsurface microbial activity associated with the IPB. The selected drilling site was Peña de Hierro on the north flank of the Río Tinto anticline. The hydrothermal activity in the area is recorded as complex-massive sulfide lenses or stockwork veins of pyrite and quartz, which occur at the upper part of the IPB volcanic sequence [22]. Three boreholes, BH1, BH4 and BH8, were continuously cored by rotary diamond-bit drilling, producing 60-mm diameter cores protected by a plastic liner. Water, with NaBr as a chemical tracer for controlling contamination, was used as the drilling fluid to refrigerate the bit. Upon retrieval, cores were flushed with N_2, sealed and transported to a nearby laboratory for geomicrobiological analysis. Samples were prepared aseptically in anaerobic conditions.

After drilling, wells were cased with PVC tubes set in clean gravel packing. Underground sampling for water and gas aquifer analysis was done by the installation of multilevel diffusion samplers (MLDS) at different depths. Ion and metal concentrations and dissolved gases were determined by ion and gas chromatography [69,70].

The groundwater entering the ore body at Peña de Hierro was characterized by analyzing springs upslope. The water from these springs is aerobic, with neutral pH and low ionic strength. The environment within the ore body was sampled by drilling boreholes BH4 and BH8 (Figure 3). Both wells reached a depth of 165 m. The water table was encountered nearly 90 m below the surface.

Rock leachate analyses were performed to detect drilling contamination and to estimate resources available to microorganisms from the solid phase. Sulfate was abundant and a good indicator of the degree of oxidation of the sulfides. Nitrite and nitrate were present in many samples. Both ferrous and ferric iron could be leached from powdered ore samples, indicating the existence of an operative iron cycle [69,70].

Borehole fluids from the MLDS were analyzed as a proxy for formation fluids. The measured pH was *ca.* 3.5 and remained acidic for two sampling years after drilling. The dissolved ferric-to-ferrous iron ratio varied along the wells ranging from 0.3 to 4.3. Sulfate concentrations were constant and lower than in rock leachates. Surprisingly, dissolved methane was detected in many MLDS samples, indicating active methanogenic activity within the ore body. The dissolved H_2 concentration averaged 25 ppm, except in the zone within the massive pyrites, just below the water table, where concentrations between 100 ppm and 1000 ppm were detected [69,70].

Figure 3. MARTE (Mars Astrobiology Research and Technology Experiment) project, borehole (BH) BH4 drilling site.

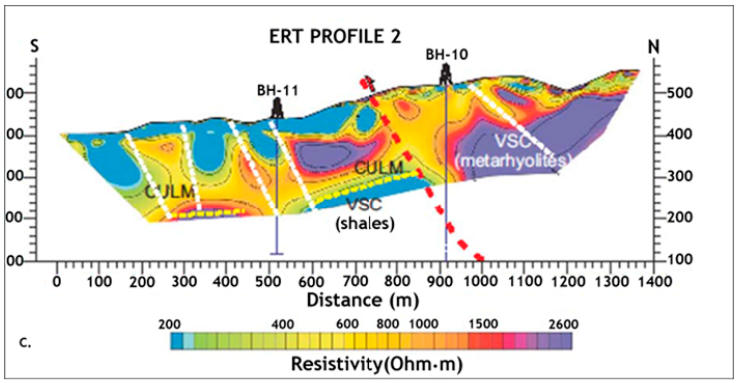
Microorganisms were detected in different uncontaminated samples using both culture-dependent and culture-independent methods. Aerobic chemolithoautotrophs, mainly pyrite and iron oxidizers, anaerobic thiosulfate oxidizers using nitrate as the electron acceptor, sulfate reducers (SRB) and methanogens were detected in enrichment cultures from core samples at different depths. Samples from both boreholes were analyzed with the microsensor, LDChip200, an antibody microarray containing 200 antibodies with different and complementary specificity, and an oligonucleotide hybridization microarray, which gave positive signals for Gram-positive bacteria, sulfur and metal oxidizers, SRBs, as well as methanogens. Hydrogenotrophic and denitrifying bacteria were also identified by 16SrRNA cloning and sequencing. Using fluorescence *in situ* hybridization (CARD-FISH), it was possible to prove the presence of active microorganisms in different uncontaminated samples [70].

The environment down-gradient from the ore body was sampled by drilling borehole BH1. Sulfate and iron concentrations were lower in the leachates from BH1 shales than those from BH4 and BH8 pyrites, while dissolved sulfate in groundwater was in much higher concentrations than in groundwater from BH4 and BH8, indicating that these waters had experienced more interaction with the sulfides of the IPB. Dissolved H_2 concentrations were lower than in BH4 and BH8, but still sufficient to make it available as a microbial electron donor. Methane concentrations were several orders of magnitude higher than at BH4 and BH8. Enrichment cultures showed mainly sulfate reducing and methanogenic activities along this borehole [69,70].

To further investigate the characteristics of the subsurface geomicrobiology of the IPB, researchers at the Centro de Astrobiología applied for an ERC project, which was granted in 2011 and which is currently being carried out (Iberian Pyrite Belt Subsurface Life Detection, IPBSL). Electric resistivity tomography (ERT) (Figure 4) and time-domain electromagnetic sounding (TDEM) were used to detect the most probable subsurface areas hosting microbial activity in deep regions of the IPB. After this geophysical information was analyzed, two wells, BH10 and BH11, with depths of 620 and 340 meters, respectively, were drilled [57].

The IPBSL drilling was performed in conditions similar to those described for the MARTE project. In addition to the geological core login in the drilling site, selected samples were obtained for mineralogical (XRD), elemental analysis (ICP-MS) and stable isotopic and petrographic analyses. The mineralogical results showed the presence of pyrite and its alteration products, such as hematite and magnesite, in both boreholes. The elemental analysis of the leachates from the core samples showed the presence of iron and other metals at different depths. The stable isotopic analysis of pyrites showed ^{34}S fractionation at different depths, which is a clear indication of sulfate reducing activities along the borehole. Furthermore, fractionation in ^{13}C was observed in samples from both boreholes, which is also a clear biosignature of microbial activity at these depths.

Figure 4. Electric resistivity tomography profile of Peña de Hierro showing the location of the selected drilling sites.



Rock leachates were analyzed by ion chromatography to determine the concentration of water soluble anions. The chromatograms showed the presence of reduced organic anions, like acetate, and oxidized inorganic inions, such as nitrate and sulfate. Protein and sugar content were also detected at different depths, indicating the presence of extant or recent microbiological activities. H_2, CO_2 and CH_4 were detected by gas chromatography of different core samples from both boreholes. Samples along the BH10 borehole were analyzed with the immunosensor, LDCChip450, an antibody microarray containing, in this case, 450 antibodies. Positive signals were detected with specific antibodies against methanogenic Archaea and SRB, which agree with the results obtained by enrichment cultures. DNA and RNA have been successfully extracted from different BH10 and BH11 samples. Most of them rendered positive PCR amplifications of bacterial and archaeal 16S rRNA gene, which are currently under analysis by cloning and massive sequencing. Samples from the two

boreholes are being analyzed by fluorescence *in situ* hybridization analysis (CARD-FISH). The results obtained so far showed positive hybridizations signals for both Bacteria and Archaea at different depths. The observed microorganisms appeared in colonies attached to mineral particles in most samples. Further hybridizations with probes selected or designed after identification of putative organisms along the boreholes are under development.

Anaerobic enrichment cultures were prepared in an anaerobic chamber using mineral salt medium with the addition of different electron donors and electron acceptors. The following activities have been detected unambiguously after more than one year of incubation using samples from both boreholes: methanogens, methanotrophs, sulfate reducers, iron oxidizers, iron reducers, denitrifiers and acetogens. From all of the available data, different hot spots have been identified. The detection of hot spots is required for the selection of samples for metagenomic and retro-transcriptomic analyses, which is under way.

The results obtained so far in both drilling projects clearly show that as groundwater enters the volcanogenic-hosted massive sulfide system of the IPB, abiotic and biological processes are activated. Electron donors available for microbial metabolism include ferrous iron, metal sulfides and H_2. Identified electron acceptors include nitrate, sulfate, ferric iron and CO_2. These compounds support a community of different microbial metabolisms. In contrast to conventional ARD models, oxidants to drive the system are supplied by the rock matrix. Only mobilization of these sources by ground water is required to promote microbial metabolisms. These observations confirmed the hypothesis that microorganisms are active in the subsurface of the IPB and are responsible for the characteristic extreme conditions detected in the Tinto basin.

6. Rio Tinto as a Mars Terrestrial Analog

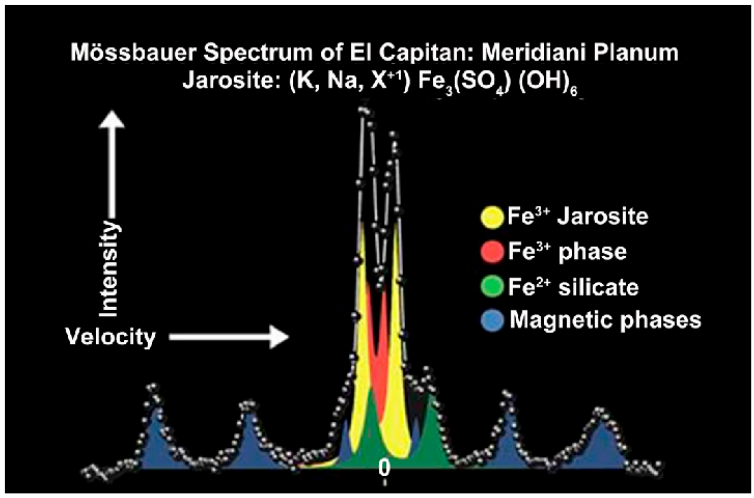
The discovery of some Noachian layered sulfate minerals at different Mars locations suggest a past aqueous, acidic, sulfate-rich environment [71–75] that might have originated from the weathering of sulfide-rich minerals [76,77]. In 2004, the NASA rover, Opportunity, from the Mars Exploration Rovers (MER) mission began its exploration on the Martian surface at Meridiani Planum. The scientific motivation for this selection was the identification by orbital observation of regionally distributed hematite, inferred to have formed under aqueous conditions on the early Martian surface [78,79]. Several hypothesis for hematite deposition have been proposed [78,80], but the observations made by the rover, Opportunity, decisively tipped the balance towards models that invoke pervasive chemical weathering of basalts and subsequent formation of hematite-rich spherules within sulfate rich sediments [71,81,82]. Jarosite, a ferric iron sulfate-hydroxide mineral identified by Mössbauer spectroscopy (Figure 5) [72], placed a particular constraint on the paleo-environmental interpretation of Meridiani outcrop rocks, as this mineral is considered to precipitate under acidic conditions [83].

As has been described previously, jarosite, goethite and hematite are iron minerals that can be found in the Tinto basin as a result of the microbial metabolism of chemolithotrophic microorganisms thriving in the high concentration of iron sulfides of the IPB. Río Tinto as a geochemical and mineralogical terrestrial Mars analogue provides an interesting perspective for the interpretation of Martian data for two reasons. First, both iron oxides and ferric sulfates are generated at Río Tinto

under well-characterized physico-chemical and biological conditions [20,36,48,50]. Second, the modern drainage, where depositional processes can be observed in action, is complemented by a historical record of deposition preserved as diagenetically stabilized sedimentary rock in terraces at different levels above the river [47,48]. The combination of ancient and modern deposits facilitates comparison with Meridiani Planum and other iron minerals regions, like Valles Marineris, Mawrth Vallis and Syrtis Major [72–75], where depositional and diagenetic processes must be inferred from ancient sedimentary rocks [84].

The mineralogical composition and sedimentary geomicrobiology reported for the Tinto basin are of use in addressing several issues of interest in Mars exploration: (i) What biological, chemical and physical processes left an interpretable record in Río Tinto rocks? (ii) How did different processes modify the initial mineralogical and chemical composition of iron-bearing precipitates, and what are the consequences of these processes for the retention of environmental or biological proxy records in Río Tinto rocks? (iii) How much of the informative proxy record might be captured using the suites of instruments used in ongoing Mars exploration missions (Mars Exploration Rovers (MER), Mars Express (MEX), Mars Reconnaissance Orbiter (MOR), Mars Science Laboratory (MSL)) or planned future missions [48,50].

Figure 5. Mössbauer spectrum of jarosite at Meridiani Planum (Courtesy of NASA/JPL-Caltech).



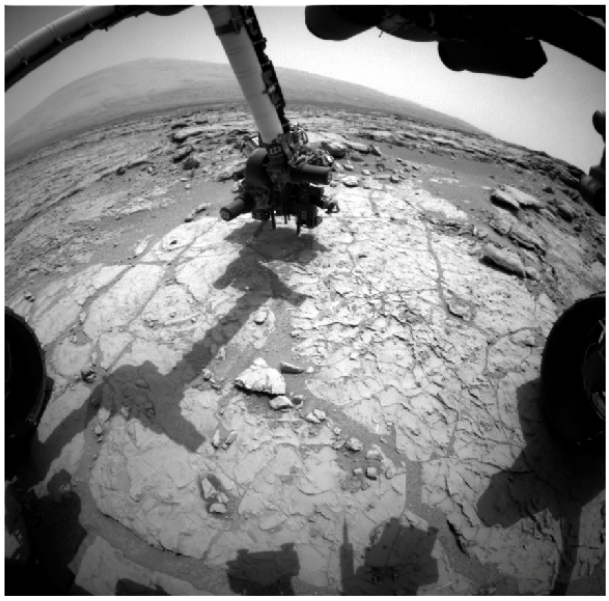
The Río Tinto and Meridiani depositional systems have both similarities and differences. The most obvious difference concerns physical setting. Whereas the Meridiani rocks accumulated via eolic and aqueous processes in an arid environment [85], Río Tinto sediments formed in seasonally arid stream beds. Moreover, the Tinto basin precipitates owe their genesis to the oxidation of hydrothermally emplaced pyrite ores. While this process has also been proposed as a source of sulfuric acid in Mars, it is not the only possible source of acidic fluids that interact with the basaltic rocks in the Meridiani region [84].

The Río Tinto Mössbauer spectra are very similar to those from Meridiani [48] and provide a useful geochemical and mineralogical analog for the processes at play when Meridian rocks were formed.

Río Tinto rocks also document the effects of diagenesis, making them doubly useful when compared to Meridiani sediments. Like Río Tinto, Martian outcrop rocks on the Meridiani plain contain hematite, but unlike the Río Tinto terraces, Meridiani outcrops remain sulfate-rich, including ferric sulfates that do not persist into the rock record at Río Tinto. This suggests that mineral formation and diagenesis occurred on Mars under extremely limited water conditions [48,80,85].

Diverse microorganisms thrive in acidic and strongly oxidizing environments, which, from an astrobiological perspective, are inferred to be at least broadly similar to those at Meridiani and other Mars iron-rich regions, at the time when their sulfate-rich sedimentary successions were deposited [73–75]. Thus, Río Tinto has helped frame some of the biological expectations of Mars exploration. The Río Tinto sedimentary deposits record aspects of the physical, chemical and biological environment of the regional ecosystem, and these indicators persist through diagenesis to provide a geochemical and geomicrobiological chronicle of the Río Tinto processes through time. By inference, Martian outcrops carry a similar potential to preserve a record of the environment and life (if it ever existed or currently exists). To improve Mars characterization, the rover, Curiosity, from the Mars Science Laboratory (MSL) mission (Figure 6) has an XRD for mineral identification and a mass spectrometer to obtain isotopic data.

Figure 6. Suite of instruments in Curiosity's mobile arm (Courtesy NASA/JPL-Caltech).

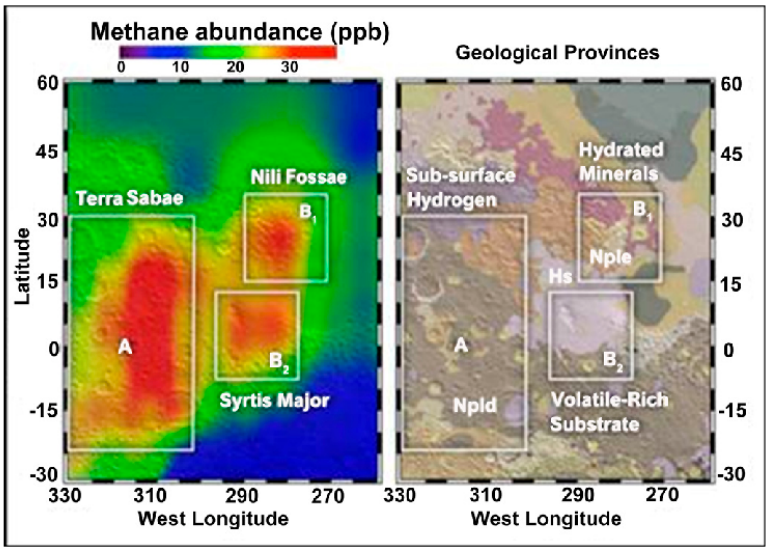


The results, to date, in the characterization of the IPB suggest that these Martian systems could support subsurface life, even if surface conditions preclude it. We have found that subsurface microbial metabolism coupled with sulfide weathering can produce large amounts of methane, which has been proposed as an atmospheric indicator of extant life on Mars (Figure 7) [86,87].

Although methane can be generated abiotically, more than 80% of Earth's methane is biologically produced as a final product of the degradation of organic matter by methanogenic Archaea.

Methanogens, with few exceptions [88,89], are generally found in habitats that share two physicochemical properties: reduced redox potentials and circumneutral pH. These conditions are the opposite of the extreme acidic and oxidative conditions existing in Rio Tinto.

Figure 7. Methane detection on Mars [81] (Courtesy of NASA/Goddard).



After the detection of methane in the Martian atmosphere [86], a systematic survey for methanogenic activity was initiated in the anoxic sediments of the Tinto basin [90]. The first site where methane production was detected was Campo de Galdierias, in the origin area of the Tinto basin. Sediments from this site showed specific positions with negative redox potential, while in the surrounding sediments, the redox potential values were highly positive and similar to those measured in the water column of the river. Microcosms were established using reduced sediments from this site spiked with different methanogenic substrates. In all cases, the production of methane was associated with a decrease in redox potentials to negative values and an increase of pH. The highest methane production was observed in microcosms spiked with methanol [90].

A second site, JL Dam, was selected to have access to deeper and more reliable sediments. Cores from this site showed well-defined black bands with negative reduced redox potentials and higher pH values than the adjacent reddish-brown sediments. Total DNA from these black bands was extracted, amplified, and sequences corresponding to *Methanosaeta concilii* were retrieved [90]. To further explore the methanogenic diversity of the cores, enrichment cultures were designed using different substrates. The highest CH_4 production occurred in the presence of a mixture of reduced organic compounds. Only *M. concilii* was detected in this microcosm, suggesting that this was the predominant methanogen in environments exposed to organic substrates. *Methanobacterium bryantii* and *Methanosarcina barkeri* were identified in cultures enriched only with H_2 or methanol, respectively [90].

As mentioned, the bulk environmental conditions at Rio Tinto are far from the conditions required to develop methanogenic Archaea, but this apparent contradiction can be resolved at the microscopic level. The generation of micro-niches might facilitate the growth of microorganisms with different

requirements from those found in the harsh bulk conditions existing in the environment. These micro-niches could be easily generated in a semi-solid matrix, such as sediments, or in a solid matrix within a subsurface rock.

The presence of methanogens in an environment controlled by oxidized iron and sulfate has important implications for the characterization of Martian methane [86,87]. The argument that the environmental conditions on Mars are not suitable for methanogenesis could be challenged by the methane production observed in the sediments of Río Tinto or the subsurface of the IPB. Considering the short life span of atmospheric methane on Mars [87], there is a possibility that extant methanogens are currently active on the subsurface of the red planet. As mentioned, the Curiosity rover, currently exploring crater Gale, is equipped with a mass spectrometer capable of measuring the carbon fractionation signal of Martian methane, which might allow its biological or abiotic origin to be clarified.

Although the history of iron in the Earth's biosphere is still an open question, we would like to suggest that the Tinto ecosystem, as well as other iron-rich acidic environments, are relics of an ancient (Archaean) iron world [9], which is probably operating in other planetary systems, such as Mars [48,50,58]. Obviously the actual conditions in which the Tinto ecosystem operates are different from the ones that may have prevailed during the Archaean or might prevail on Mars, but the properties of the microorganisms isolated and characterized so far in this environment allow us to extrapolate their performance in these systems.

Liquid water seems to be an absolute requirement for life. As indicated, liquid water is abundant in the Tinto basin, both on the surface and underground. Conversely, due to environmental constraints, water appears only in solid or vapor phases on the current Mars surface [91]. Climatic studies of the early atmospheric evolution of Mars [92] indicate that during the Noachian, the atmospheric pressure was high enough to sustain substantial amounts of liquid water on its surface, explaining the above mentioned water-related features. Although we have the orbital technology to reveal the possible existence of liquid water on the subsurface of Mars, so far, there is only indirect evidence of widespread subterranean ice [93], its quantification in the polar water-ice [94] and the characterization of brines from the polar region [95]. However, images from Mars, as well as spectral data provided by different instruments in orbit and on the surface of the planet support the existence of distinctive episodes of water on Mars' surface in the past [71,96–103]. Although there is only a remote possibility that the Martian iron mineral formations are the product of chemolithoautotrophy, the microbial diversity found in the Tinto basin, with metabolisms compatible with the conditions prevailing on Mars, suggest that microorganisms may have grown or are still growing in places where mineral and water converge. Obviously, Río Tinto is not the only acidic environment of astrobiological interest. In recent years, diverse extreme acidic environments have been identified as terrestrial analogues of Mars, including the seasonally dry acidic lakes of Kalgoorly in Australia [104,105], the cold acid drainage systems in the Canadian Artic [106–109] and the ARD in King George Island in Antarctica (Figure 8) [110]. Most of these environments have been analyzed from a geological and mineralogical point of view. A thorough geomicrobiological characterization of these sites and others to be explored will eventually complement the information already obtained in the IPB. From a Martian perspective, it would be very useful to take into consideration the possible existence of micro-niches by designing a drilling mission to gather information on the existence of redox gradients of possible use by chemolithotrophic microorganisms.

Figure 8. Iron bioformations along the coastline of Cardozo Cove at King George Island [110].

7. Conclusions

Forty years ago, the Viking mission, considered the first astrobiological mission devoted to the search for signs of life on Mars, concluded that life had little chance of developing there due to the extreme conditions detected on its surface. Since then, important advances in microbiology, especially in the characterization of extreme environments, have challenged this pessimistic point of view. Acidophiles are of special interest, because they form the only known natural extreme environment generated by the metabolic activity of chemolithoautotrophic microorganisms. The characterization of the Río Tinto basin, an extreme acidic environment, has addressed some basic issues, including the origin of the extreme conditions of the habitat, the identification and isolation of the microorganisms responsible for these conditions and the existence of micro-niches in the sediments and the subsurface that facilitate the development of microorganisms with requirements incompatible with the bulk conditions existing in the environment. The discovery of some Noachian iron lithological units on Mars, similar to those produced biologically in the Tinto basin, gave Río Tinto the status of a geochemical and mineralogical terrestrial analogue that enables us to better understand those geomicrobiological processes that may have driven the generation of iron oxides and sulfates on Mars.

Acknowledgments

This revision has been supported by the ERC project ERC250350-IPBSL.

Author Contributions

Ricardo Amils and David Fernández-Remolar wrote the paper, and the IPBSL team has participated in the development of the IPBSL drilling project. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Des Marais, D.J.; Nuth, J.A.; Allamandola, L.J.; Boss, A.P.; Farmer, J.D.; Hoehler, T.M.; Jakosky, B.M.; Meadows, V.S.; Pohorille, A.; Runnegar, B.; *et al.* The NASA Astrobiology Roadmap. *Astrobiology* **2008**, *8*, 715–730.
2. Margulis, L.; Mazur, P.; Barghoorn, E.S.; Halvorson, H.O.; Jukes, T.H.; Kaplan, I.R. The Viking Mission: Implications for life in the Vallis Marineris area. *J. Mol. Evol*. **1979**, *14*, 223–232.
3. Johnson, D.B.; Hallberg, K.B. The microbiology of acidic mine waters. *Res. Microbiol*. **2003**, *154*, 466–473.
4. Ehrlich, H.L.; Newman, D.K. *Geomicrobiology*, 5th ed.; CRC Press: Boca Raton, FL, USA, 2008.
5. Sand, W.; Gehrke, T.; Hallma, R.; Schippers, A. Sulfur chemistry, biofilm and the (in)direct attack mechanisms—A critical evaluation of bacterial leaching. *Appl. Microbiol. Biotechnol.* **1995**, *43*, 961–966.
6. Sand, W.; Gehrke, T.; Jozsa, P.G.; Schippers, A. Biochemistry of bacterial leaching—Direct *vs.* indirect bioleaching. *Hydrometall* **2001**, *59*, 159–175.
7. Colmer, A.R.; Temple, K.L.; Hinkle, H.E. An iron-oxidizing bacterium from the acid drainage of some bituminous coal mines. *J. Bacteriol*. **1950**, *59*, 317–328.
8. Pronk, J.T.; Bruyn, J.C.; Bos, P.; Kuenen, J.G. Anaerobic growth of Thiobacillus ferrooxidans. *Appl. Environ. Microbiol*. **1992**, *58*, 2227–2230.
9. Amils, R.; González-Toril, E.; Gómez, F.; Fernández-Remolar, D.; Rodríguez, N.; Malki, M.; Zuluaga, J.; Aguilera, A.; Amaral-Zettler, L.A. Importance of chemolithotrophy for early life on Earth: The Tinto River (Iberian Pyritic Belt) case. In *Origins*; Seckbach, J., Ed.; Springer: Amsterdam, The Netherlands, 2004; pp. 463–480.
10. Rawlings, D.E. Heavy metal mining using microbes. *Annu. Rev. Microbiol*. **2002**, *56*, 65–91.
11. Benz, M.; Brune, A.; Schink, B. Anaerobic and aerobic oxidation of ferrous iron at neutral pH by chemoheterotrophic nitrate-reducing bacteria. *Arch. Microbiol*. **1998**, *169*, 159–165.
12. Widdel, F.; Schnell, S.; Heising, S.; Ehrenreich, A.; Assmus, B.; Schink, B. Ferrous iron oxidation by anoxygenic phototrophic bacteria. *Nature* **1993**, *162*, 834–836.
13. Godd, T. The deep hot biosphere. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6045–6049.
14. Bachofen, R.; Ferloni, P.; Flynn, L. Microorganisms in the subsurface. *Microbiol. Res*. **1998**, *153*, 1–22.
15. Pedersen, K. Exploration of deep intraterrestrial microbial life: Current perspectives. *FEMS Microbiol. Lett*. **2000**, *185*, 9–16.
16. Chapelle, F.H.; O'Nelly, K.; Bradley, P.M.; Methé, B.A.; Ciufo, S.A.; Knobel, L.L.; Lovley, D.R. A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* **2002**, *415*, 312–315.

17. Leblanc, M.; Morales, J.A.; Borrego, J.; Elbaz-Poulichet, F. 4500-year-old mining pollution in Southwestern Spain: Long-Term implications for modern mining pollution. *Econ. Geol.* **2000**, *95*, 655–662.
18. López-Archilla, A.I.; Marín, I.; Amils, R. Microbial community composition and ecology of an acidic aquatic environment: The Tinto River, Spain. *Microbiol. Ecol.* **2001**, *41*, 20–35.
19. Amaral-Zettler, L.A.; Gómez, F.; Zettler, E.; Keenan, B.G.; Amils, R.; Sogin, M.L. Microbiology: Eukaryotic diversity in Spain's River of Fire. *Nature* **2002**, *417*, 137, doi:10.1038/417137a.
20. González-Toril, E.; LLobet-Brosa, E.; Casamayor, E.O.; Amann, R.; Amils, R. Microbial ecology of an extreme acidic environment, the Tinto River. *Appl. Environ. Microbiol.* **2003**, *69*, 4853–4865.
21. Boulter, C.A. Did both extensional tectonics and magmas act as major drivers of convection cells during the formation of the Iberian Pyrite Belt massive sulphide deposits? *J. Geol. Soc. London* **1996**, *153*, 181–184.
22. Leistel, J.M.; Marcoux, E.; Theiblemont, D.; Quesada, C.; Sánchez, A.; Almodóvar, G.R.; Pascual, E.; Saez, R. The volcanic-hosted massive sulphide deposits of the Iberian Pyrite Belt. *Miner. Depos.* **1997**, *33*, 2–30.
23. Lescuyer, J.L.; Leistel, J.M.; Mrcoux, E.; Milési, J.P.; Thiéblemont, D. Late Devonian-Early Carboniferous peak sulphide mineralization in the Western Hercynides. *Miner. Depos.* **1997**, *33*, 208–220.
24. González-Toril, E.; Aguilera, A.; Rodríguez, N.; Fernández-Remolar, D.; Gómez, F.; Díaz, E.; García-Moyano, A.; Sanz, J.L.; Amils, R. Microbial ecology of Río Tinto, a natural extreme acidic environment of biohydrometallurgical interest. *Hydrometall* **2010**, *104*, 329–333.
25. García-Moyano, A.; González-Toril, E.; Aguilera, A.; Amils, R. Prokaryotic community composition and ecology of macroscopic floating filaments from an extreme acidic environment, Río Tinto, (SW, Spain). *Syst. Appl. Microbiol.* **2007**, *30*, 601–614.
26. García-Moyano, A.; González-Toril, E.; Aguilera, A.; Amils, R. Comparative microbial ecology study of the sediments and the water column of the Río Tinto, an extreme acidic environment. *FEMS Microbiol. Ecol.* **2012**, *81*, 303–314.
27. Ohmura, N.; Sasaki, K.; Matsumoto, N.; Saiki, H. Anaerobic respiration using Fe^{3+}, S^0 and H_2 in the chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans*. *J. Bacteriol.* **2002**, *18*, 2081–2087.
28. Malki, M.; González-Toril, E.; Sanz, J.L.; Gómez, F.; Rodríguez, N.; Amils, R. Importance of the iron cycle in biohydrometallurgy. *Hydrometall* **2006**, *83*, 223–228.
29. Coupland, K.; Johnson, D.B. Evidence that the potential for dissimilatory ferric iron reduction is widespread among acidophilic heterotrophic bacteria. *FEMS Microbiol. Lett.* **2008**, *279*, 30–35.
30. Malki, M.; de Lacey, A.L.; Rodríguez, N.; Amils, R.; Fernández, V.M. Preferential use of an anode as an electron acceptor by an acidophilic bacterium in the presence of oxygen. *Appl. Environ. Microbiol.* **2008**, *74*, 4472–4476.
31. Sánchez-Andrea, I.; Rodríguez, N.; Amils, R.; Sanz, J.L. Microbial diversity in anaerobic sediments at Río Tinto, a naturally acidic environment with high heavy metal content. *Appl. Environ. Microbiol.* **2011**, *77*, 6085–6093, doi:10.1128/AEM.00654-11.

32. Sánchez-Andrea, I.; Rojas-Ojeda, P.; Amils, R.; Sanz, J.L. Screening of anaerobic activities in sediments of an acidic environment: Tinto River. *Extremophiles* **2012**, *16*, 829–839.
33. Sánchez-Andrea, I.; Stams, A.J.M.; Amils, R.; Sanz, J.L. Enrichment and isolation of acidophilic sulfate-reducing bacteria from Tinto River sediments. *Environ. Microbiol. Rep.* **2013**, *5*, 672–678, doi:10.1111/1758-2229.12066.
34. Lu, S.; Gischkat, S.; Reiche, M.; Akob, D.M.; Hallberg, K.B.; Küsel, K. Ecophysiology of Fe-cycling bacteria in acidic sediments. *Appl. Environ. Microbiol.* **2010**, *76*, 8174–8183.
35. Hallberg, K.B.; Johnson, D.B. Biodiversity of acidophilic prokaryotes. *Adv. Appl. Microbiol.* **2001**, *49*, 37–84.
36. González-Toril, E.; Gómez, F.; Malki, M.; Amils, R. The Isolation and Study of Acidophilic Microorganisms. In *Extremophiles, Methods in Microbiology*; Rainey, F.A., Oren, A., Eds.; Elsevier Academic Press: San Diego, CA, USA, 2006; Volume 35, pp. 471–510.
37. Rawlings, D.E. Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microb. Cell Factor.* **2005**, *4*, 13–28.
38. Aguilera, A.; Manrubia, S.C.; Gómez, F.; Rodríguez, N.; Amils, R. Eukaryotic community distribution and its relationship to water physicochemical parameters in an extreme acidic environment, Río Tinto (Southwestern Spain). *Appl. Environ. Microbiol.* **2006**, *72*, 5325–5330.
39. Aguilera, A.; Zettler, E.; Gómez, F.; Amaral-Zettler, L.; Rodríguez, N.; Amils, R. Distribution and seasonal variability in the benthic eukaryotic community of Río Tinto (SW, Spain), and acidic, high metal extreme environment. *Syst. Appl. Microbiol.* **2007**, *30*, 531–546.
40. Amaral-Zettler, L.; Zettler, E.R.; Theroux, S.M.; Palacios, C.; Aguilera, A.; Amils, R. Microbial community structure across the tree of life in the extreme Río Tinto. *ISME J.* **2010**, *5*, 42–50, doi:10.1038/ismej.2010.101.
41. López-Archilla, A.I.; González, A.E.; Terrón, M.C.; Amils, R. Diversity and ecological relationships of the fungal populations of an acidic river of Southwestern Spain: The Tinto River. *Can. J. Microbiol.* **2005**, *50*, 923–934.
42. Rodríguez, N.; Menéndez, N.; Tornero, J.; Amils, R.; de la Fuente, V. Internal iron biomineralization in *Imperata cylindrica*, a perennial grass: Chemical composition, speciation and plant localization. *New Phytol.* **2005**, *165*, 781–789.
43. De la Fuente, V.; Rufo, L.; Rodríguez, N.; Amils, R.; Zuluaga, J. Metal accumulation screening of the Río Tinto flora (Huelva, Spain). *Biol. Trace Elem. Res.* **2010**, *134*, 318–341.
44. Franco, A.; Rufo, L.; Rodríguez, N.; Amils, R.; de la Fuente, V. Iron absorption, localization and biomineralization of *Cynodon. dactylon*, a perennial grass from the Río Tinto basin (SW Iberian Peninsula). *J. Plant Nutr. Soil Sci.* **2013**, *176*, 836–842.
45. Amils, R.; de la Fuente, V.; Rodríguez, N.; Zuluaga, J.; Menéndez, N.; Tornero, J. Composition, speciation and distribution of iron minerals in *Imperata cylindrica*. *Plant Physiol. Biochem.* **2007**, *45*, 335–340.
46. De la Fuente, V.; Rodríguez, N.; Amils, R. Immunocytochemical analysis of the subcellular distribution of ferritin in *Imperata cylindrica* (L.) Raeuschel, an iron hyperaccumulator plant. *Acta Histochim.* **2012**, *114*, 232–236.

47. Fernández-Remolar, D.C.; Rodríguez, N.; Gómez, F.; Amils, R. Geological record of an acidic environment driven by iron hydrochemistry: The Tinto River system. *J. Geophys. Res.* **2003**, *108*, doi:10.1029/2002JE001918.
48. Fernández-Remolar, D.C.; Morris, R.V.; Gruener, J.E.; Amils, R.; Knoll, A.H. The Rio Tinto Basin, Spain: Mineralogy, sedimentary geobiology and implications for interpretation of ourcrop rocks of meridiani Planum, Mars. *Earth Planet Sci. Lett.* **2005**, *240*, 149–167.
49. Fernández-Remolar, D.C.; Knoll, A.H. Fossilization potential of iron-bearing minerals in acidic environments of Rio Tinto, Spain: Implications for Mars exploration. *Icarus* **2008**, *194*, 72–85.
50. Amils, R.; González-Toril, E.; Fernández-Remolar, D.; Gómez, F.; Aguilera, A.; Rodríguez, N.; Malki, M.; García-Moyano, A.; González-Fairén, A.; de la Fuente, V.; *et al.* Extreme environments as Mars terrestrial analogs: The Río Tinto case. *Planet Space Sci.* **2007**, *55*, 370–381.
51. Oggerin, M.; Tornos, F.; Rodríguez, N.; del Moral, C.; Sánchez-Román, M.; Amils, R. Specific jarosite biomineralization by *Purpureocillium. lilacinum*, an acidophilic fungi isolated from Río Tinto. *Environ. Microbiol.* **2013**, *15*, 2228–2237.
52. Fernández-Remolar, D.C.; Preston, L.J.; Sánchez-Román, M.; Izawa, M.R.M.; Huang, L.; Southam, G.; Banerjee, N.R.; Osinski, G.R.; Flemming, R.; Gómez-Ortíz, D.; *et al.* Carbonate precipitation under bulk acidic conditions as a potential biosignature for searching life on Mars. *Earth Planet Sci. Lett.* **2012**, *351*, 13–26.
53. Sánchez-Román, M.; Fernández-Remolar, D.; Amils, R.; Sánchez-Navas, A.; Schmid, T.; Martín-Uriz, P.S.; Rodríguez, N.; McKenzie, J.A. Microbial mediated formation of Fe-carbonate minerals under extreme acidic conditions. *Sci. Rep.* **2014**, *4*, doi:10.1038/srep04767.
54. Colín-García, M.; Kanawati, B.; Harir, M.; Schmidt-Kopplin, P.; Amils, R.; Parro, V.; García, M.; Fernández-Remolar, D. Detection of peptidic sequences in the ancient acidic sediments of Río Tinto, Spain. *Orig. Life Evol. Biosph.* **2011**, *41*, 523–527.
55. Preston, L.; Shuster, J.; Fernández-Remolar, D.; Banerjee, N.; Osinski, G.R.; Southam, G. The preservation and degradation of filamentous bacteria and biomolecules within iron oxide deposits at Rio Tinto, Spain. *Geobiology* **2011**, *9*, 233–249.
56. Geen, A.; van Adkins, J.F.; Boyle, E.A.; Nelson, C.H.; Palanques, A. A 120-yr record of widespread contamination from mining of the Iberian Pyrite Belt. *Geology* **1997**, *25*, 291–294.
57. Gómez-Ortiz, D.; Fernández-Remolar, D.; Granda, A.; Quesada, C.; Granda, T.; Prieto-Ballesteros, O.; Molina, A.; Amils, R. Identification of the subsurface sulfide bodies responsible for acidity in Río Tinto source water, Spain. *Earth Plant. Sci. Lett.* **2014**, *391*, 36–41.
58. Fernández-Remolar, D.; Gómez, F.; Prieto-Ballesteros, O.; Schelble, R.T.; Rodríguez, N.; Amils, R. Some ecological mechanisms to generate habitability in planetary subsurface areas by chemolithotrophic communities: The Río Tinto subsurface ecosystem as a model system. *Astrobiol*ogy **2008**, 8, 157–173.
59. Moreno, C.; Capitán, M.A.; Doyle, M.; Nieto, J.M.; Ruiz, F.; Sáez, R. Edad mínima del gossan de Las Cruces: Implicaciones sobre la edad del inicio de los ecosistemas extremos en la Faja Pirítica Ibérica. *Geogaceta* **2003**, *33*, 67–70. (In Spainish)
60. Gómez, F.; Aguilera, A.; Amils, R. Soluble ferric iron as an effective protective agent against UV radiation: Implications for early life. *Icarus* **2007**, *191*, 352–359.

61. Gómez, F.; Mateo-Martí, E.; Prieto-Ballesteros, O.; Martín-Gago, J.; Amils, R. Protection of chemolithotrophic bacteria exposed to Mars environmental conditions. *Icarus* **2010**, *209*, 482–487, doi:10.1016/j.icarus.2010.05.027.
62. Archibald, F. *Lactobacillus plantarum*, an organism not requiring iron. *FEMS Microbiol. Lett.* **1983**, *19*, 29–32.
63. Martin, J.H. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleooceanography* **1990**, *5*, 1–13.
64. Boyd, W.P.; Jickells, T.; Law, C.S.; Blain, S.; Boyle, E.A.; Buesseler, K.O.; Coale, K.H.; Cullen, J.J.; de Baar, H.J.W.; Follows, M.; *et al.* Mesoscale iron enrichment experiments 1993–2005: Synthesis and future directions. *Science* **2007**, *315*, 612–617.
65. Reid, R.T.; Live, D.H.; Faulkner, D.J.; Buttler, A. A siderophore from a marine bacterium with an exceptional ferric iron affinity constant. *Nature* **1993**, *366*, 455–458.
66. Braun, V.; Killmann, H. Bacterial solution to the iron supply problems. *Trends Biochem. Sci.* **1999**, *24*, 104–109.
67. Gómez, F.; Fernández-Remolar, D.; González-Toril, E.; Amils, R. The Tinto River, an Extreme Gaian Environment. In *Scientists Debate Gaia 2000*; Margulis, L., Miller, J., Boston, P., Schneider, S., Crist, C., Eds.; MIT Press: Boston, MA, USA, 2003; pp. 321–333.
68. Anbar, A.D.; Knoll, A.H. Proterozoic ocean chemistry and evolution: A bioinorganic bridge. *Science* **2002**, *297*, 1137–1142.
69. Fernández-Remolar, D.; Prieto-Ballesteros, O.; Rodríguez, N.; Gómez, F.; Amils, R.; Gomez-Elvira, J.; Stoker, C. Underground habitats found in the Río Tinto Basin: A model for subsurface life habitats on Mars. *Astrobiol*ogy **2008**, *8*, 1023–1046.
70. Puente-Sánchez, F.; Moreno-Paz, M.; Rivas, L.A.; Cruz-Gil, P.; García-Villadangos, M.; Gómez, M.J.; Postigo, M.; Garrido, P.; González-Toril, E.; Briones, C.; *et al.* Deep subsurface sulfate reduction and methanogenesis in the Iberian Pyrite Belt revealed through geochemistry and molecular biomarkers. *Geobiology* **2014**, *12*, 34–47.
71. Squyres, S.W.; Crotzinger, J.P.; Arvidson, R.E.; Bell, J.F., III; Calvin, W.; Christensen, P.R.; Clark, B.C.; Crisp, J.A.; Farrand, W.H.; Herkenhoff, K.E.; *et al. In situ* evidence for an ancient aqueous environment in Meridiani Planum, Mars. *Science* **2004**, *306*, 1709–1714.
72. Klingelhöfer, G.; Morris, R.V.; Bernhardt, B.; Schröder, C.S.; de Souza, P.A., Jr.; Yen, A.; Gellert, R.; Evlanov, E.N.; Zubkov, B.; Foh, J.; *et al.* Jarosite and hematite at *Meridiani Planum* from the Mössbauer spectrometer on the Opportunity rover. *Science* **2005**, *306*, 1740–1745.
73. Milliken, R.E.; Swayze, G.A.; Arvidson, R.E.; Bishop, J.L.; Clark, R.N.; Ehlmann, B.L.; Green, R.O.; Grotzinger, J.P.; Morris, R.V.; Murchie, S.L.; *et al.* Opaline silica in young deposits on Mars. *Geology* **2008**, *36*, 847–850.
74. Farrand, W.H.; Glotch, T.D.; Rice, J.W.; Hurowitz, J.A.; Swayze, G. Discovery of jarosite within Mawrth Vallis region of Mars: Implications for the geological history of the region. *Icarus* **2009**, *204*, 478–488.
75. Ehdmann, B.L.; Mustard, J.F. An in-situ record of major environmental transitions on early Mars at Northeast Syrtis Major. *Geophys. Res. Lett.* **2012**, *39*, doi:10.1029/2012GL051594.
76. Fairen, A.G.; Fernández-Remolar, D.; Dohm, J.M.; Baker, V.R.; Amils, R. Inhibition of carbonate synthesis in acidic oceans from Mars. *Nature* **2004**, *431*, 423–426.

77. Zolotov, M.; Shock, E. Formation of jarosite-bearing deposits through aqueous oxidation of pyrite at the Meridiani Planum, Mars. *Geophys. Res. Lett.* **2005**, *32*, doi:10.1029/2005GL024253.
78. Christensen, P.R.; Bandfield, J.L.; Clark, R.N.; Edgett, K.S.; Hamilton, V.E.; Hoefen, T.; Kieffer, H.H.; Kuzmin, R.O.; Lane, M.D.; Malin, M.C.; *et al.* Detection of crystalline hematite mineralization on Mars by the thermal emission spectrometer evidence for near-surface water. *J. Geophys. Res.* **2000**, *104*, 9623–9642.
79. Christensen, P.R.; Morris, R.V.; Lane, M.D.; Banfield, J.L.; Malin, M.C. Global mapping of martian hematite mineral deposits: remnants of water-driven processes on early Mars. *J. Geophys. Res.* **2001**, *106*, 23873–23885.
80. Christensen, P.R.; Ruff, S.W. Formation of the hematite-bearing unit in Meridiani Planum: Evidence for deposition in standing water. *J. Geophys. Res. Planet.* **2004**, *109*, E08003.
81. Rieder, R.; Gellert, R.; Anderson, R.C.; Brückner, J.; Clark, B.C.; Dreibus, G.; Economou, T.; Klingelhöfer, G.; Lugmair, G.W.; Ming, D.W.; *et al.* Chemistry of rocks and soils at Meridiani Planum from the alpha particle X-ray spectrometer. *Science* **2004**, *306*, 1746–1749.
82. McLennan, S.M.; Bell, J.F., III; Calvin,W.M.; Christensen, P.R.; Clark, B.C.; de Souza, P.A.; Farmer, J.; Farrand, W.H.; Fike, D.A.; Gellert, R.; *et al.* Provenance and diagenesis of the Burns formation, Meridiani Planum, Mars. *Earth Planet. Sci. Lett.* **2005**, *240*, 95–121.
83. Bigham, J.M.; Schwertmann, U.; Traina, S.J.; Winland, R.L.; Wolf, M. Schwertmannite and the chemical modelling of iron in acid sulphate waters. *Geochim. Cosmochim. Acta* **1996**, *60*, 2111–2121.
84. Knoll, A.H.; Carr, M.; Clark, B.; Des Marais, D.J.; Farmer, J.D.; Fische, W.W.; Grotzinger, J.P.; McLennan, S.M.; Malin, M.; Schröder, C.; *et al.* An astrobiological perspective on Meridiani Planum. *Earth Planet. Sci. Lett.* **2005**, *240*, 179–189.
85. Grotzinger, J.P.; Arvidson, R.E.; Bell, J.F., III; Calvin, W.; Clark, B.C.; Fike, D.A.; Golombek, M.; Greeley, R.; Haldemann, A.; Herkenhoff, K.E.; *et al.* Stratigraphy, sedimentology and depositional environment of the Burns formation, Meridiani Planum, Mars. *Earth Planet. Sci. Lett.* **2005**, *240*, 11–72.
86. Formisano, V.; Atreya, S.; Encrenaz, T.; Ignatiev, N.; Giuranna, M. Detection of methane in the atmosphere of Mars. *Science* **2004**, *306*, 1758–1761.
87. Mumma, M.J.; Villanueva, G.L.; Novak, R.E.; Hewagama, T.; Bonev, B.P.; DiSanti, M.A.; Mandell, A.; Smith, M.D. Strong release of methane on Mars in Northen Summer 2003. *Science* **2009**, *323*, 1041–1045.
88. Kotsyurbenko, O.R.; Friedrich, M.W.; Simankova, M.V.; Nozhenvnikova, A.N.; Golyshin, P.N.; Timmis, K.N.; Conrad, R. Shift from acetoclastic to H_2 dependent methanogenesis in a West Siberian peat bog at low pH values and isolation of an acidophilic *Methanobacterium* strain. *Appl. Environ. Microbiol.* **2007**, *73*, 2344–2348.
89. Taconi, K.A.; Zappi, M.E.; French, W.T.; Brown, L.R. Methanogenesis under acidic pH conditions in a semi-continuous reactor system. *Bioresour. Technol.* **2008**, *99*, 8075–8081.
90. Sanz, J.L.; Rodríguez, N.; Díaz, E.; Amils, R. Methanogenesis in the sediments of Río Tinto, an extreme acidic environment. *Environ. Microbiol.* **2011**, *13*, 2336–2341.
91. Jakosky, B.M.; Haberle, R.M. The Seasonal Behavior of Water on Mars. In *Mars*; Kieffer, H.H., Ed.; University of Arizona Press: Tucson, AZ, USA, 1992; pp. 969–1016.

92. Carr, M.H. Retention of an atmosphere on early Mars. *J. Geophys. Res.* **1999**, *104*, 21897–21909.
93. Boynton, W.V.; Feldman, W.C.; Squyres, S.W.; Prettyman, T.H.; Brückner, J.; Evans, L.G.; Reedy, R.C.; Starr, R.; Arnold, J.R.; Drake, D.M.; *et al.* Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* **2002**, *297*, 81–85.
94. Bibring, J.P.; Langevin, Y.; Gendrin, A.; Gondet, B.; Poulet, F.; Berthé, M.; Soufflot, A.; Arvidson, R.; Mangold, N.; Mustard, J.; *et al.* Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science* **2005**, *307*, 1576–1581.
95. Smith, P.H.; Tamppari, L.K.; Arvidson, R.E.; Bass, D.; Blaney, D.; Boynton, W.V.; Carswell, A.; Catling, D.C.; Clark, B.C.; Duck, T.; *et al.* H_2O at the Phoenix landing site. *Science* **2009**, *325*, 58–61.
96. Parker, T.J.; Gorsline, D.S.; Saunders, R.S.; Pieri, D.C.; Schneeberger, D.M. Coastal geomorphology of the martian northern plains. *J. Geophys. Res.* **1993**, *98*, 11061–11078.
97. Malin, M.C.; Edgett, K.S. Sedimentary rocks of early Mars. *Science* **2000**, *290*, 1927–1937.
98. Malin, M.C.; Edgett, K.S. Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* **2003**, *302*, 1931–1934.
99. Fairén, A.G.; Dohm, J.M.; Baker, V.R.; de Pablo, M.A.; Ruiz, J.; Ferris, J.; Anderson, R. Episodic flood inundations of the northern plains of Mars. *Icarus* **2003**, *165*, 53–67.
100. Fairén, A.G. A cold and wet Mars. *Icarus* **2010**, *208*, 165–175.
101. Bhattacharya, J.P.; Payenberg, T.H.D.; Lang, S.C.; Bourke, M. Dynamic river channels suggest a long-lived Noachian crater lake on Mars. *Geophys. Res. Lett.* **2005**, *32*, doi:10.1029/2005GL022747.
102. Poulet, F.; Bibring, J.P.; Mustard, J.F.; Gendrin, A.; Mangold, N.; Langevin, Y.; Arvidson, R.E.; Gondet, B.; Gómez, C.; the Omega Team. Phyllosilicates on Mars and implications for early martian climate. *Nature* **2005**, *438*, 623–627.
103. Heldmann, J.L.; Toon, O.B.; Pollard, W.H.; Mellon, M.T.; Pitlick, J.; McKay, C.P.; Andersen, D.T. Formation of Martian gullies by the action of liquid water flowing under current Martian environmental conditions. *J. Geophys. Res.* **2005**, *110*, doi:10.1029/2004JE002261.
104. Baldridge, A.M.; Hook, S.J.; Crowley, J.K.; Marion, G.M.; Kargel, J.S.; Michalski, J.L.; Thomson, B.J.; de Souza Filho, C.R.; Bridges, N.T.; Brown, A.J.; *et al.* Contemporaneous deposition of phyllosilicates and sulfates: Using Australin acidic lake deposits to describe geochemical variability on Mars. *Geophys. Res. Lett.* **2009**, *36*, doi:10.1029/2009GL040069.
105. Benison, K.O.; LaClair, D.A. Modern an ancient extremely acid saline deposits: Terrestrial analogs for martian environments. *Astrobiology* **2003**, *3*, 609–618.
106. Michel, F.A.; Everdingen, R.O. Formation of jarosite deposits on Cretaceous shales in the Fort Norman area, Northwest Territories. *Can. Mineral.* **1987**, *25*, 221–226.
107. Lacelle, D.; Levillé, R. Acid drainage generation and associated Ca-Fe-SO_4 minerals in a periglacial environment, Eagle Plains, Northen Yukon, Canada: A potential analogue for low-temperature sulfate formation on Mars. *Planet. Space Sci.* **2010**, *5*, 509–521.
108. Batller, M.W.; Osinki, G.R.; Lim, D.S.S.; Dávila, A.F.; Michel, F.A.; Craig, M.A.; Izawa, M.R.M.; Leoni, L.; Slater, G.F.; Fairén, A.G.; *et al.* Characterization of the acidic cold seep emplaced jarosite: Golden Deposit, NWT, Canada, as an analogue for jarosite deposition on Mars. *Icarus* **2012**, *242*, 382–398.

109. West, L.; McGovern, D.J.; Onston, T.C.; Morris, R.V.; Suchecki, P.; Pratt, L.M. High Lake gossan deposit: An Artic analogue for ancient Martian surficial processes? *Plant. Sapce. Sci.* **2009**, *57*, 1302–1311.
110. Dold, B.; González-Toril, E.; Aguilera, A.; López-Pamo, E.; Cisternas, M.E.; Bucchi, F.; Amils, R. Acid Rock Drainage and Rock Weathering in Antarctica: important sources for iron cycling in the Southern Ocean. *Environ. Sci. Technol.* **2013**, *47*, 6129–6136, doi:10.1021/es305141b.

Tessaracoccus lapidicaptus sp. nov., an actinobacterium isolated from the deep subsurface of the Iberian Pyrite Belt

Tessaracoccus lapidicaptus sp. nov., an actinobacterium isolated from the deep subsurface of the Iberian pyrite belt

Fernando Puente-Sánchez,[1] Mónica Sánchez-Román,[2] Ricardo Amils[2,3] and Víctor Parro[1]

Correspondence
Fernando Puente-Sánchez
puentesf@cab.inta-csic.es

[1]Departments of Molecular Evolution, Centro de Astrobiología (INTA-CSIC), Carretera de Ajalvir Km 4, Torrejón de Ardoz, 28850 Madrid, Spain

[2]Planetology and Habitability, Centro de Astrobiología (INTA-CSIC), Carretera de Ajalvir Km 4, Torrejón de Ardoz, 28850 Madrid, Spain

[3]Centro de Biología Molecular 'Severo Ochoa' (UAM-CSIC), Cantoblanco, 28049 Madrid, Spain

A novel actinobacterium, designated IPBSL-7^T, was isolated from a drilling core 297 m deep obtained from the Iberian Pyrite Belt. The strain was isolated anaerobically using nitrate as the electron acceptor. 16S rRNA gene sequence analysis revealed that it was related to *Tessaracoccus flavescens* SST-39^T (95.7 % similarity), *Tessaracoccus bendigoensis* Ben 106^T (95.7 %), *Tessaracoccus lubricantis* KSS-17Se^T (95.6 %) and *Tessaracoccus oleiagri* SL014B-20A1^T (95.0 %), while its similarity to any other member of the family *Propionibacteriaceae* was less than 94 %. Cells were non-motile, non-spore-forming, Gram-positive, oval to rod-shaped, and often appeared in pairs or small groups. The strain was facultatively anaerobic, oxidase-negative, catalase-positive and capable of reducing nitrate. Colonies were circular, convex, smooth and colourless. The organism could grow at between 15 and 40 °C, with an optimal growth at 37 °C. The pH range for growth was from pH 6 to 9, with pH 8 being the optimal value. Strain IPBSL-7^T had peptidoglycan type A3-γ', with LL-diaminopimelic acid as the diagnostic diamino-acid and glycine at position 1 of the peptide subunit. The dominant menaquinone was MK-9(H$_4$) (93.8 %). The major cellular fatty acid was anteiso-C$_{15:0}$ (55.0 %). The DNA G+C content was 70.3 mol%. On the basis of phenotypic and phylogenetic results, strain IPBSL-7^T can be differentiated from previously described species of the genus *Tessaracoccus* and, therefore, represents a novel species, for which the name *Tessaracoccus lapidicaptus* sp. nov. is proposed. The type strain is IPBSL-7^T (=CECT 8385^T=DSM 27266^T).

The genus *Tessaracoccus*, belonging to the family *Propionibacteriaceae*, was first described by Maszenan *et al.* (1999) and contains four species with validly published names at the time of writing: *Tessaracoccus bendigoensis* (Maszenan *et al.*, 1999), *Tessaracoccus flavescens* (Lee & Lee, 2008), *Tessaracoccus lubricantis* (Kämpfer *et al.*, 2009) and *Tessaracoccus oleiagri* (Cai *et al.*, 2011). All four species are Gram-positive and non-motile actinobacteria. The cell-wall peptidoglycan contains LL-diaminopimelic acid (LL-DAP). MK-9(H4) is the major menaquinone and anteiso-C$_{15:0}$ is the predominant cellular fatty acid.

Abbreviation: LL-DAP, LL-diaminopimelic acid.

The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain IPBSL-7^T is KF668596.

Three supplementary figures are available with the online version of this paper.

One gram of powder sample from a core sampled at 297 m depth and stored under anaerobic conditions was aseptically removed in a clean N_2-filled glovebox and used to inoculate a 100 ml vial of anoxic F4 medium (0.4 g NaCl l^{-1}; 0.4 g NH_4Cl l^{-1}; 0.3 g $MgCl_2 . 6H_2O$ l^{-1}; 0.05 g $CaCl_2 . 2H_2O$ l^{-1}; 1 g yeast extract l^{-1}; 2 g peptone l^{-1}; 1 g glucose l^{-1}; 1 g succinic anhydride l^{-1}; 7.5 g $NaHCO_3$ l^{-1}; 0.5 g KH_2PO_4 l^{-1}; −1; 0.5 g Na_2S l^{-1}; 1 mg resazurin l^{-1}). A negative control was performed by keeping a second vial open inside the glovebox for the duration of the inoculation procedure. The two vials were sealed with a gastight rubber septum and an aluminium cap, taken out of the glovebox and incubated at 30 °C. Growth was followed by visual inspection and microscopic examination. After 1 week, growth was detected in the inoculated vial and 1 ml of the culture was transferred to a 100 ml vial of fresh anoxic modified F4 medium (similar to F4 medium but with no peptone and 0.1 g yeast extract l^{-1}) and incubated at 30 °C for another week. Pure cultures were

obtained by inoculating different dilutions of this second enrichment culture onto plates of modified F4 medium supplemented with 20 g agar l^{-1} in a clean N_2-filled glovebox and incubating them at 30 °C in an anaerobic jar. After finding that strain IPBSL-7^T was capable of aerobic growth, cells were stored on tryptone soy broth (TSB; BD-Difco) in 20 % (v/v) glycerol at −80 °C.

Cells grown in liquid TSB for 8 h at 30 °C and 200 r.p.m. were used to examine cell morphology and motility with a light microscope (BX40; Olympus). Cell morphology was also examined with a JEOL JEM-1010 transmission electron microscope (Fig. S1, available in the online Supplementary Material) at the Centro Nacional de Microscopía Electrónica (CNME, Madrid, Spain), after preparing the samples as described elsewhere (Ferrero *et al.*, 2013), but without adding 5 % $K_3Fe(CN)_6$ in the post-fixation step. Colony morphology and pigmentation were observed after incubation on tryptic soy agar (TSA; BD-Difco) for 5 days at 30 °C. The 16S rRNA gene was amplified by colony PCR using the universal primers 27f and 1492r (Lane, 1991), with the following PCR conditions: 95 °C, 5 min; (95 °C, 30 s; 48 °C, 45 s; 72 °C, 1.5 min) × 30 cycles; 72 °C, 10 min. The amplicon was sequenced with an ABI PRISM BigDye Terminator cycle sequencing kit (Applied Biosystems) and an automatic DNA sequencer (model 3730xl; Applied Biosystems). The resulting 16S rRNA gene sequence had a length of 1434 nt, and was aligned with the SILVA SSU Ref NR 99 reference database (Pruesse *et al.*, 2007) by using SINA (Pruesse *et al.*, 2012). The ARB software package (Ludwig *et al.*, 2004) was used to identify closely related bacteria and to reconstruct phylogenetic trees. Candidate sequences for tree reconstruction were aligned by using CLUSTAL W (Larkin *et al.*, 2007). The neighbour-joining (Saitou & Nei, 1987) with Felsenstein correction (Felsenstein, 1985a) and maximum-parsimony (Fitch, 1971) algorithms included in the ARB package were used for phylogenetic inference (Figs 1 and S2). The robustness of the reconstructed trees was evaluated by bootstrap analysis (Felsenstein, 1985b) of 1000 resampled datasets. The 16S rRNA gene sequence similarities of strain IPBSL-7^T with *T. flavescens* SST-39^T, *T. bendigoensis* Ben 106^T, *T. lubricantis* KSS-17Se^T and *T. oleiagri* SL014B-20A1^T were 95.7, 95.7, 95.6 and 95.0 % respectively. The novel strain also showed 93.8 % 16S rRNA sequence similarity with *Brooklawnia cerclae* DSM 19609^T (Bae *et al.*, 2006), which did not belong to the *Tessaracoccus* cluster. The 16S rRNA gene sequence similarities of strain IPBSL-7^T with other type species of the family *Propionibacteriaceae* were less than 93.2 %.

Results of the physiological characterization are given in the species description. Growth response to temperature was tested in liquid TSB medium at 8, 15, 25, 30, 37 and 40 °C. Growth response to pH was tested in liquid TSB medium, buffered with a 10 mM citrate buffer for pH 5 and 5.5, a 10 mM potassium phosphate buffer for pH 6, 6.5, 7, 7.5 and 8, and a 10 mM HEPES buffer for pH 8.5 and 9. Salt tolerance was tested in liquid TSB medium supplemented with 1, 2, 4 or 8 % (w/v) NaCl. Bacterial growth was determined by following OD_{600} variations of duplicate cultures shaken at 200 r.p.m. for 2 days and comparing them with uninoculated controls. The effects of temperature, pH and salt concentration on bacterial growth were also determined for *T. bendigoensis* Ben 106^T (type species of the genus *Tessaracoccus*) and *T. flavescens* SST-39^T (closest phylogenetic relative of strain IPBSL-7 ^{T}), by using the same methodology described above. The tolerance to heavy metals was determined by growing serial dilutions of the strain on TSA plates amended with different concentrations of $NiSO_4$, $ZnSO_4$, $CuSO_4$, $CdSO_4$ or $CoSO_4$. The final metal ion concentration varied from 0.5 to 20 mM. Other physiological properties (substrate utilization, enzymic activities, tolerance to antibiotics and other compounds) of the strain were examined with commercial test systems [GPIII (Biolog), API20A, API20NE, APIZym (bioMérieux)] and antibiotic discs by the DSMZ Identification Service

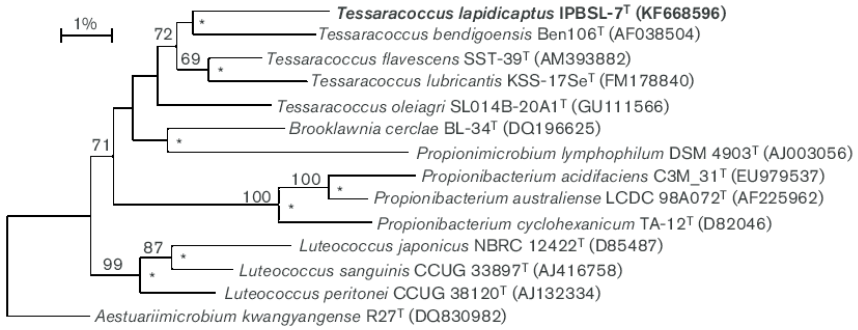


Fig. 1. Neighbour-joining tree based on 16S rRNA gene sequences showing the relationship of strain IPBSL-7^T with other members of the family *Propionibacteriaceae*. Bootstrap values smaller than 65 % are not shown. Nodes marked with asterisks were also found in the maximum-parsimony tree. Bar, 1 nt substitution per 100 nt.

Table 1. Differential characteristics between strain IPBSL-7^T and its most closely related phylogenetic neighbours

Strains: 1, IPBSL-7^T; 2, *T. bendigoensis* Ben 106^T; 3, *T. flavescens* SST-39^T. All strains are Gram-positive, facultatively anaerobic and contain LL-DAP in the cell-wall peptidoglycan. All strains are positive for nitrate reduction, aesculin hydrolysis, glucose fermentation, glycerol fermentation and mannose fermentation. All strains are positive for the following enzymic activities: esterase, esterase lipase, leucine arylamidase, α-galactosidase, β-galactosidase, α-glucosidase and β-glucosidase. All strains are negative for indole production, glucose fermentation, arginine hydrolysis, urease activity and gelatin hydrolysis. All strains were unable to grow on the following substrates: 3-*O*-methyl D-glucose, D-fucose, L-fucose, L-rhamnose, inosine, D-glucose 6-phosphate, D-fructose 6-phosphate, D-aspartic acid, D-serine, gelatin, glycyl-L-proline, L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L-histidine, L-pyroglutamic acid, L-serine, pectin, D-galacturonic acid, L-galacturonic acid γ-lactone, mucic acid, quinic acid, D-saccharic acid, *p*-hydroxyphenylacetic acid, methyl pyruvate, citric acid, α-ketoglutaric acid, D-malic acid, L-malic acid, bromosuccinic acid, Tween 40, γ-amino-n-butyric acid, β-hydroxybutyric acid, α-ketobutyric acid, acetoacetic acid, propionic acid, acetic acid and sodium formate. All strains were negative for the following enzymic activities: alkaline phosphatase, lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, acid phosphatase, β-glucuronidase, α-mannosidase and α-fucosidase. +, Positive; −, negative; W, weakly positive.

Characteristic	1	2	3
Cell morphology	Oval to rods (0.45 × 0.5–1.0 μm)	Cocci (0.5–1.1 μm)*	Rods (0.6 × 1.2 μm)*
Temperature for growth (range, optimum; °C)	15–40, 37	15–40, 25	15–40, 30
pH for growth (range, optimum)	6–9, 8	5–9, 8	6–9, 8.5
NaCl concentration for growth (range, optimum; %)	0–2, 0	0–8, 4	0–8, 0
Carbon source utilization			
Dextrin	−	−	+
Maltose	+	−	+
Trehalose	+	−	+
Cellobiose	+	−	+
β-Gentiobiose	−	−	+
Sucrose	+	−	+
Turanose	+	−	+
Stachyose	+	−	+
Raffinose	+	+	−
α-Lactose	+	−	+
Melibiose	+	+	+
Methyl β-D-glucoside	−	−	+
D-Salicin	+	−	+
N-Acetyl-D-glucosamine	−	−	+
N-Acetyl-β-D-mannosamine	−	−	+
N-Acetyl-D-galactosamine	−	−	+
N-Acetyl-neuraminic acid	−	−	+
D-Glucose	+	−	+
D-Mannose	+	−	+
D-Fructose	+	−	+
D-Galactose	+	−	+
L-Fucose	−	−	+
L-Rhamnose	−	−	+
Inosine	−	−	+
D-Sorbitol	+	−	+
D-Mannitol	+	−	−
D-Arabitol	+	−	+
myo-Inositol	+	−	+
Glycerol	+	−	+
Pectin	−	−	+
D-Gluconic acid	−	−	+
D-Glucuronic acid	−	−	+
Glucuronamide	−	−	+
D-Lactic acid methyl ester	−	−	+
L-Lactic acid	−	−	+
α-Hydroxybutyric acid	−	−	+

Table 1. cont.

Characteristic	1	2	3
Acetoacetic acid	+	−	+
Acid production from:			
D-Mannitol	W	+	−
Lactose	−	+	−
Sucrose	−	+	W
Maltose	−	+	+
D-Xylose	−	+	+
L-Arabinose	−	+	+
Enzyme activities			
Naphthol-AS-BI-phosphohydrolase	−	W	W
N-Acetyl-β-glucosaminidase	−	W	+
Chemotaxonomic data			
Major menaquinones	MK-9(H_4) (93.8 %), MK-9(H_6) (4.1 %), MK-9(H_2) (2.1 %)	MK-9(H_4)*	MK-9(H_4) (84 %), MK-8(H_0) (12 %), MK-7(H_2) (4 %)*
Phospholipids†	PGL, PG, GL, PI, PL, L	DPG, PG, PI, PL*	DPG, PG*
DNA G+C content (mol%)	70.3	74*	68.4*

*Data taken from Cai *et al.* (2011).
†PGL, unknown phosphoglycolipid; DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; GL, unknown glycolipid; PI, phosphatidylinositol; PL, unknown phospholipid; L, unknown polar lipid.

(Braunschweig, Germany). All commercial tests were performed according to the manufacturers' instructions on cells grown in TSB at 30 °C.

The commercial tests listed above and the cellular fatty acid analyses were also performed on *T. bendigoensis* Ben 106^T and *T. flavescens* SST-39^T cells grown on TSB at 30 °C.

Results of chemotaxonomic analyses are given in the species description. Chemotaxonomic characterization (polar lipids, cellular fatty acids, respiratory quinones and peptidoglycan structure) was carried out by the DSMZ Identification Service. Polar lipids and respiratory quinones were extracted from 100 mg freeze-dried cell material using the two-stage method described by Tindall (1990a, b). Respiratory quinones were extracted using methanol/hexane, followed by phase separation into hexane. Respiratory lipoquinones were separated by TLC on silica gel (art. no. 805 023; Macherey-Nagel), using hexane/tert-butylmethylether (9 : 1, v/v) as solvent. UV absorbing bands corresponding to the different quinone classes were removed from the plate and further analysed by HPLC on a LDC Analytical (Thermo Separation Products) HPLC fitted with a reverse-phase column (2 mm × 125 mm, 3 µm, RP18; Macherey-Nagel) using methanol/heptane 9 : 1 (v/v) as the eluant. Respiratory lipoquinones were detected at 269 nm. Polar lipids were separated by two-dimensional silica gel TLC (art. no. 818 135; Macherey-Nagel). The first direction was developed in chloroform/methanol/water (65 : 25 : 4, by vol.), and the second in chloroform/methanol/acetic acid/water (80 : 12 : 15 : 4, by vol.). Total lipid material was detected using molybdatophosphoric acid and functional groups were detected using specific spray reagents, Full details are given by Tindall *et al.* (2007).

Fatty acid methyl esters were obtained and analysed as described elsewhere (Miller, 1982; Kuykendall *et al.*, 1988; Kämpfer & Kroppenstedt, 1996). Peptidoglycan was isolated, purified and analysed according to published protocols (Schumann, 2011). DNA G+C content analysis was carried out by the DSMZ Identification Service according to published protocols (Cashion *et al.*, 1977; Tamaoka & Komagata, 1984; Mesbah *et al.*, 1989).

Strain IPBSL-7^T was resistant to the following antibiotics (µg of antibiotic per disc, unless otherwise specified): colistin (10), nitrofurantoin (100), pipemidic acid (20), fosfomycin (50) and nystatin (100 units); weakly sensitive to aztreonam (30), gentamicin (10), amikacin (30), norfloxacin (10), polymyxin B (300 units) and kanamycin (30); and sensitive to penicillin G (10 units), oxacillin (5), ampicillin (10), ticarcillin (75), meziocillin (30), cefalotin (30), cefazolin (30), cefotaxime (30), imipenem (10), tetracycline (30), chloramphenicol (30), vancomycin (30), erythromycin (15), lincomycin (15), ofloxacin (5), bacitracin (10 units), neomycin (30), doxycyclin (30), ceftriaxone (30), clindamycin (10), moxifloxacin (5), linezolid (30), quinupristin/dalfopristin (15), teicoplanin (30) and piperacillin/tazobactam (40).

Phylogenetic analyses showed that strain IPBSL-7^T clustered together with other strains from the genus *Tessaracoccus* (Fig. 1, Fig. S2). Additionally, the strain was positive for nitrate reduction and its peptidoglycan type was A3γ' (Schleifer & Seidl, 1985). These two characteristics are shared among all

Table 2. Fatty acid composition of strain IPBSL-7^T and the most closely related strains

Strains: 1, IPBSL-7^T; 2, *T. bendigoensis* Ben 106^T; 3, *T. flavescens* SST-39^T. Values are percentages of total fatty acids.

Fatty acid	1	2	3
Saturated fatty acids			
$C_{12:0}$	0.8	0.3	–
$C_{14:0}$	3.3	2.48	1.4
$C_{14:0}$ 2-OH	–	2.0	–
$C_{15:0}$	–	0.6	–
$C_{15:0}$ 2-OH	–	0.3	–
$C_{16:0}$	5.0	9.5	3.1
$C_{17:0}$	2.7	–	–
$C_{18:0}$	1.0	0.4	0.9
Unsaturated fatty acids			
$C_{13:1}$ at 12-13	–	1.3	–
$C_{16:1}\omega9c$	1.1	–	–
$C_{17:1}\omega6c$	0.8	–	–
$C_{17:1}\omega8c$	2.6	–	–
$C_{18:1}\omega7c$	–	0.4	–
$C_{18:1}\omega9c$	1.4	1.7	–
$C_{20:4}\omega6,9,12,15c$	–	1.4	–
Branched fatty acids			
iso-$C_{14:0}$	3.6	2.5	9.6
iso-$C_{14:0}$ 3-OH	–	0.8	0.7
iso-$C_{15:0}$	7.5	2.5	14.2
iso-$C_{16:0}$	6.4	1.2	14.7
iso-$C_{17:0}$	1.9	–	1.8
iso-$C_{18:0}$	0.5	–	1.1
anteiso-$C_{13:0}$	–	0.5	–
anteiso-$C_{15:0}$	55.0	48.8	46.9
anteiso-$C_{15:1}$ A	0.5	–	0.6
anteiso-$C_{17:0}$	2.8	0.5	1.2
anteiso-$C_{17:1}\omega9c$	0.8	–	–
Summed features*			
1	–	2.7	–
2	–	0.8	–
3	2.5	–	–
4	–	11.27	3.0
5	–	0.5	–

*Summed features are groups of fatty acids that could not be separated by GC with the MIDI system. Summed features 1, 2, 3, 4 and 5 comprised $C_{13:0}$ 3-OH/$C_{15:1}$, $C_{16:1}\omega7c$/ iso-$C_{15:0}$ 2-OH, $C_{16:1}\omega6c$/$C_{16:1}\omega7c$, iso-$C_{17:1}$ I/anteiso-$C_{17:1}$ B and $C_{18:2}\omega6,9c$/anteiso-$C_{18:0}$, respectively.

known members of the genus *Tessaracoccus*, but are absent in closely related genera such as *Brooklawnia* (Bae *et al.*, 2006) and *Propionimicrobium* (Stackebrandt *et al.*, 2002). This further supports the inclusion of strain IPBSL-7^T in the genus *Tessaracoccus*.

The main morphological, physiological and chemotaxonomic characteristics that differentiate strain IPBSL-7^T from other related strains representing the genus *Tessaracoccus* are shown in Tables 1 and 2. There were differences in salt tolerance, substrate utilization, enzymic activities and membrane composition. Interestingly, diphosphatidylglycerol, which is present in the previously described species of the genus *Tessaracoccus*, was not detected in strain IPBSL-7^T (Fig. S3). This feature could be diagnostic for distinguishing strain IPBSL-7^T from other related strains. On the basis of these phenotypic and phylogenetic results, strain IPBSL-7^T should be included in the genus *Tessaracoccus* as a representative of a novel species, for which the name *Tessaracoccus lapidicaptus* sp. nov. is proposed.

Description of *Tessaracoccus lapidicaptus* sp. nov.

Tessaracoccus lapidicaptus (la.pi.di.cap′tus. L. n. *lapis lapidis* stone; L. adj *captus* captive; N.L. masc. adj. *lapidicaptus* stone-captive).

Facultatively anaerobic, Gram-positive, oxidase-negative, catalase-positive, non-endospore-forming, non-motile,

oval to rod-shaped cells (0.45 × 0.5–1.0 μm) that often occur in pairs or small groups (Fig. S1). Colonies grown on TSA are circular, convex, smooth and colourless. Growth occurs between 15 and 40 °C, with an optimum at 37 °C. The pH range for growth is from pH 6 to 9, with pH 8 being the optimal value. Tolerates up to 2 % (w/v) NaCl, 1 mM Zn^{2+}, 2 mM Ni^{2+}, 1 mM Cu^{2+} and 1 mM Co^{2+}. Is able to grow in the presence of guanidine hydrochloride, nalidixic acid, potassium tellurite and sodium butyrate; but not in the presence of 1 % sodium lactate, fusidic acid, D-serine, niaproof 4, tetrazolium violet, tetrazolium blue, lithium bromide or sodium bromate. Utilizes maltose, trehalose, cellobiose, sucrose, turanose, stachyose, raffinose, α-lactose, melibiose, D-salicin, α-D-glucose, D-mannose, D-fructose, D-galactose, D-sorbitol, D-mannitol, D-arabitol, *myo*-inositol, glycerol and acetoacetic acid. Cannot use dextrin, gentiobiose, *N*-acetyl-D-glucosamine, *N*-acetyl-β-D-mannosamine, *N*-acetyl-D-galactosamine, *N*-acetyl-D-neuraminic acid, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, inosine, D-glucose 6-phosphate, D-fructose 6-phosphate, D-aspartic acid, D-serine, gelatin, glycyl-L-proline, L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L-histidine, L-pyroglutamic acid, L-serine, pectin, D-galacturonic acid, D-galacturonic acid lactone, D-gluconic acid, D-glucuronic acid, glucuronamide, mucic acid, quinic acid, D-saccharic acid, *p*-hydroxyphenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, citric acid, α-ketoglutaric acid, D-malic acid, L-malic acid, bromosuccinic acid, Tween 40, γ-aminobutyric acid, α-hydroxybutyric acid, β-hydroxy-DL-butyric acid, α-ketobutyric acid, propionic acid, acetic acid and formic acid. Produces acid from the following substrates: D-glucose, D-mannitol, glycerol, cellobiose, D-mannose, raffinose, D-rhamnose. Hydrolyses aesculin, but not gelatin. Is not able to produce indole from L-tryptophan and is urease-negative. No anaerobic acid production is detected from the following substrates: lactose, sucrose, maltose, salicin, D-xylose, D-arabinose, melezitose, D-sorbitol, trehalose. Positive for the following enzymic activities: esterase, esterase lipase, leucine arylamidase, α-galactosidase, β-galactosidase, α-glucosidase and β-glucosidase. No activity from the following enzymes was detected: alkaline phosphatase, lipase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsine, acid phosphatase, naphthol-AS-BI-phosphohydrolase, β-glucuronidase, *N*-acetyl-β-glucosaminidase, α-manosidase and α-fucosidase. The peptidoglycan was of the A3γ′ type and contained LL-DAP. The polar lipid composition includes phosphatidylglycerol, phosphatidylinositol, one phosphoglycolipid, two different phospholipids, three glycolipids and an unidentified polar lipid (Fig. S3). The predominant cellular fatty acid is anteiso-$C_{15:0}$, but significant amounts of iso-$C_{15:0}$, iso-$C_{16:0}$ and $C_{16:0}$ are also found. More detailed results can be found in Table 2. The major respiratory quinones are MK-9(H_4), MK-9(H_6) and MK-9(H_2).

The type strain, IPBSL-7^T (=CECT 8385^T=DSM 27266^T), was isolated from a 297-metre-depth drilling core obtained from the Iberian Pyrite Belt. The DNA G + C content of the type strain is 70.3 mol%.

Acknowledgements

We thank Patxi San Martín for useful advice and comments, and Miriam García-Villadangos and Marina Postigo-Cacho for technical assistance. We also would like to thank the personnel from the DSMZ identification service and the CNME microscopy service. The present work was supported by the European Research Council (ERC) Advanced Grant no. 250350 entitled 'Iberian Pyrite Belt Subsurface Life Detection' (IPBSL), and by the Subdirección General de Proyectos de Investigación of the Spanish Ministerio de Economía y Competitividad (MINECO) grant no. AYA2011-24803. F. P.-S. has a JAE-pre fellowship from the Spanish Consejo Superior de Investigaciones Científicas (CSIC). There are no conflicts of interests.

References

Bae, H. S., Moe, W. M., Yan, J., Tiago, I., da Costa, M. S. & Rainey, F. A. (2006). *Brooklawnia cerclae* gen. nov., sp. nov., a propionate-forming bacterium isolated from chlorosolvent-contaminated groundwater. *Int J Syst Evol Microbiol* **56**, 1977–1983.

Cai, M., Wang, L., Cai, H., Li, Y., Wang, Y. N., Tang, Y. Q. & Wu, X. L. (2011). *Salinarimonas ramus* sp. nov. and *Tessaracoccus oleiagri* sp. nov., isolated from a crude oil-contaminated saline soil. *Int J Syst Evol Microbiol* **61**, 1767–1775.

Cashion, P., Holder-Franklin, M. A., McCully, J. & Franklin, M. (1977). A rapid method for the base ratio determination of bacterial DNA. *Anal Biochem* **81**, 461–466.

Felsenstein, J. (1985a). Phylogenies and the comparative method. *Am Nat* **125**, 1–15.

Felsenstein, J. (1985b). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Ferrero, J. J., Alvarez, A. M., Ramírez-Franco, J., Godino, M. C., Bartolomé-Martín, D., Aguado, C., Torres, M., Luján, R., Ciruela, F. & Sánchez-Prieto, J. (2013). β-Adrenergic receptors activate exchange protein directly activated by cAMP (Epac), translocate Munc13-1, and enhance the Rab3A-RIM1α interaction to potentiate glutamate release at cerebrocortical nerve terminals. *J Biol Chem* **288**, 31370–31385.

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool* **20**, 406–416.

Kämpfer, P. & Kroppenstedt, R. M. (1996). Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Can J Microbiol* **42**, 989–1005.

Kämpfer, P., Lodders, N., Warfolomeow, I. & Busse, H. J. (2009). *Tessaracoccus lubricantis* sp. nov., isolated from a metalworking fluid. *Int J Syst Evol Microbiol* **59**, 1545–1549.

Kuykendall, L. D., Roy, M. A., O'Neill, J. J. & Devine, T. E. (1988). Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradorhizobium japonicum*. *Int J Syst Bacteriol* **38**, 358–361.

Lane, D. J. (1991). 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*, pp. 115–175. Edited by E. Stackebrandt & M. Goodfellow. New York: Wiley.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2. *Bioinformatics* **23**, 2947–2948.

Lee, D. W. & Lee, S. D. (2008). *Tessaracoccus flavescens* sp. nov., isolated from marine sediment. *Int J Syst Evol Microbiol* **58**, 785–789.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S. & other authors (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363–1371.

Maszenan, A. M., Seviour, R. J., Patel, B. K., Schumann, P. & Rees, G. N. (1999). *Tessaracoccus bendigoensis* gen. nov., sp. nov., a Gram-positive coccus occurring in regular packages or tetrads, isolated from activated sludge biomass. *Int J Syst Bacteriol* **49**, 459–468.

Mesbah, M., Premachandran, U. & Whitman, W. (1989). Precise measurement of the G+C content of deoxyribonucleic acid by high performance liquid chromatography. *Int J Syst Bacteriol* **39**, 159–167.

Miller, L. T. (1982). Single derivatization method for bacterial fatty acid methyl esters including hydroxy acids. *J Clin Microbiol* **16**, 584–586.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**, 7188–7196.

Pruesse, E., Peplies, J. & Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Schleifer, K. H. & Seidl, P. H. (1985). Chemical composition and structure of murein. In *Chemical Methods in Bacterial Systematics*, pp. 201–219. Edited by M. Goodfellow & D. E. Minnikin. London: Academic Press.

Schumann, P. (2011). Peptidoglycan structure. *Methods Microbiol.* **38**, 101–129.

Stackebrandt, E., Schumann, P., Schaal, K. P. & Weiss, N. (2002). *Propionimicrobium* gen. nov., a new genus to accommodate *Propionibacterium lymphophilum* (Torrey 1916) Johnson and Cummins 1972, 1057^{AL} as *Propionimicrobium lymphophilum* comb. nov. *Int J Syst Evol Microbiol* **52**, 1925–1927.

Tamaoka, J. & Komagata, K. (1984). Determination of DNA base composition by reversed-phase high-performance liquid chromatography. *FEMS Microbiol Lett* **25**, 125–128.

Tindall, B. J. (1990a). A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. *Syst Appl Microbiol* **13**, 128–130.

Tindall, B. J. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol Lett* **66**, 199–202.

Tindall, B. J., Sikorski, J., Smibert, R. M. & Krieg, N. R. (2007). Phenotypic characterization and the principles of comparative systematics. In *Methods for General and Molecular Microbiology*, 3rd edn, pp. 330–393. Edited by C. A. Reddy, T. J. Beveridge, J. A. Breznak, G. Marzluf, T. M. Schmidt & L. R. Snyder. Washington, DC: American Society for Microbiology.

Nucleation of Fe-Rich phosphates and carbonates on microbial cells and exopolymeric substances

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 September 2015
doi: 10.3389/fmicb.2015.01024

Nucleation of Fe-rich phosphates and carbonates on microbial cells and exopolymeric substances

Mónica Sánchez-Román[1]*, *Fernando Puente-Sánchez*[2], *Víctor Parro*[2] and *Ricardo Amils*[1,3]

[1] Department of Planetology and Habitability, Centro de Astrobiología (INTA-CSIC), Madrid, Spain, [2] Department of Molecular Evolution, Centro de Astrobiología (INTA-CSIC), Madrid, Spain, [3] Department of Virology and Microbiology, Centro de Biología Molecular Severo Ochoa, Madrid, Spain

Although phosphate and carbonate are important constituents in ancient and modern environments, it is not yet clear their biogeochemical relationships and their mechanisms of formation. Microbially mediated carbonate formation has been widely studied whereas little is known about the formation of phosphate minerals. Here we report that a new bacterial strain, *Tessarococcus lapidicaptus,* isolated from the subsurface of Rio Tinto basin (Huelva, SW Spain), is capable of precipitating Fe-rich phosphate and carbonate minerals. We observed morphological differences between phosphate and carbonate, which may help us to recognize these minerals in terrestrial and extraterrestrial environments. Finally, considering the scarcity and the unequal distribution and preservation patterns of phosphate and carbonates, respectively, in the geological record and the biomineralization process that produces those minerals, we propose a hypothesis for the lack of Fe-phosphates in natural environments and ancient rocks.

Keywords: microbial, bacterial precipitates, nanoglobules, Tessarococcus, vivianite, siderite

OPEN ACCESS

Edited by:
Maria Dittrich,
University of Toronto, Canada

Reviewed by:
Javier Alvaro,
Centre of Astrobiology, Spain
Mohamed Larbi Merroun,
University of Granada, Spain
Muriel Pacton,
Université Lyon 1, France

***Correspondence:**
Mónica Sánchez-Román,
Department of Planetology and Habitability, Centro de Astrobiología (INTA-CSIC), Ctra. Ajalvir. Km 4, Torrejón de Ardoz, 28850 Madrid, Spain
sanchezromanmonica@gmail.com

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *29 April 2015*
Accepted: *26 August 2015*
Published: *22 September 2015*

Citation:
Sánchez-Román M, Puente-Sánchez F, Parro V and Amils R (2015) Nucleation of Fe-rich phosphates and carbonates on microbial cells and exopolymeric substances. Front. Microbiol. 6:1024. doi: 10.3389/fmicb.2015.01024

Introduction

Authigenic ferrous iron-rich minerals like vivianite [$Fe_3(PO_4)_2 \times 8H_2O$] and siderite ($Fe_2CO_3$) are used as indicators of paleoenvironmental conditions, diagenetic evolution of sedimentary sequences (Last and De Deckker, 1990; Manning et al., 1999; Sapota et al., 2006) and biosignatures (Vuillemin et al., 2013; Sánchez-Román et al., 2014). They are usually found associated in organic rich environments like lacustrine (Lemos et al., 2007; Rothe et al., 2014) and deep-sea sediments (Dijkstra et al., 2014), swamps, sewage, and wastewater treatment plants (Postma, 1981; Lovley et al., 1991). Vivianite is considered the most important sink of phosphorus in reducing natural systems, being a significant parameter controlling the trophic status of lakes (Nriagu and Dell, 1974; Manning et al., 1991). Therefore, it can exert significant controls over the geochemical cycles of P and Fe (Veeramani et al., 2011) in reducing sediments in which iron and phosphorous are highly mobile and the sulfide ion is not produced in high concentration (Manning et al., 1991). On the other hand, the majority of the carbonate minerals on Earth surface are of biogenic origin (Moore, 1989; Riding, 2006) and the process of carbonate precipitation can be the most important factor controlling the global carbon cycling (Ridgwell and Zeebe, 2005; Dupraz et al., 2009). Vivianite and siderite usually occur associated with pyrite (FeS_2) in veins of copper, tin, iron, and gold ores (Craig and Vaughan, 1994; Wiberg et al., 2001). Furthermore, these two iron-rich minerals are used as iron fertilizer (Eynard et al., 1992; Rakshit et al., 2008; Sánchez-Alcalá et al., 2012) and

more rarely vivianite has been used as phosphorous fertilizer (Mikhailov, 1940; Nelipa, 1961). In addition, vivianite is also found in decaying plants and animal tissues, bones, shells, anthropogenic compounds, human wastes, and archeological settings (Jakobsen, 1988; McGowan and Prangnell, 2006; Nutt and Swihart, 2012).

Vivianite is a significant mineral because links forensic medicine (Thali et al., 2011), physical anthropology (McGowan and Prangnell, 2006), biology and climate geology (Sapota et al., 2006). On the other hand, siderite is also an important carbonate mineral that provides information about past climatic events on Earth and Mars (Ellwood et al., 1998; Fairen et al., 2004; Tomkinson et al., 2013). Actually, siderite and vivianite are significant constituents of martian meteorites and Mars surface (Valley et al., 1997; Dyar et al., 2014). In contrast to vivianite, siderite is found in much greater abundance in ancient rocks than in modern environments (Ohmoto et al., 2004; Kholodov and Butuzova, 2008). The formation of these two iron rich minerals is generally attributed to the activity of iron reducing bacteria (Mortimer and Coleman, 1997; Orange et al., 2009; Lee et al., 2010), and they have significant implications for microbial metabolism in sediments (Fredrickson et al., 1998). It is known that siderite can be formed within the sub-oxic, sulfate-reducion and methanogenic biogeochemical zones within the sediment column (Wilkinson et al., 2000).

Recently, Fe-rich sulfide (pyrite), sulfate [jarosite, $KFe_3(OH)_6(SO_4)_2$], and carbonate (siderite) minerals have been found in Rio Tinto basin (Fernández-Remolar et al., 2012) and their formation have been also related to iron-reducing fungi and bacteria (Oggerin et al., 2013; Sánchez-Román et al., 2014). Those minerals together with vivianite are known as important minerals in the iron biogeochemical cycle (Raiswell and Canfield, 2012). Rio Tinto is an acidic system in which microorganisms play an important role by determining the speciation of iron and can also cause considerable iron accumulation through biomineralization (Fernández-Remolar et al., 2012; Oggerin et al., 2013; Sánchez-Román et al., 2014). Furthermore, this acid-sulfate system enriched in iron is considered one of the potential analogs for early life on Earth and Mars (Fernández-Remolar et al., 2012). In order to better understand the nucleation and formation processes of iron carbonate and phosphate minerals, here, we present for the first time microbially mediated primary precipitation of siderite and vivianite in anaerobic culture experiments under Earth's surface conditions using a bacterial strain, *Tessarococcus lapidicaptus*, isolated from the subsurface of Rio Tinto (Puente-Sánchez et al., 2014a). The nucleation, chemical composition, texture and morphology of the bioprecipitates have been studied using a combination of high resolution transmission electron microscopy (TEM), scanning electron microscopy (SEM), sensitive energy dispersive X-ray Spectroscopy (EDS), and X-ray powder diffraction (XRD). We demonstrate that *T. lapidicaptus* produces spatially restricted supersaturated conditions and can overcome kinetic barrier to nucleate phosphate and carbonate nanocrystals in its cells and secreted EPS, respectively. We propose that microbial nanostructures, nanocrystals, and crystalline nanoparticles, are not related to a single microbial group or to a specific microbial metabolism but to a wide range of microorganisms. Finally, we discuss the mechanism of formation of both phosphate and carbonate and their significance and implication in natural systems.

Materials and Methods

Microorganism

Tessaracoccus lapidicaptus CECT 8385 (= DSM 27266) is a gram-positive, non-spore forming, oval to rod shaped, nitrate-reducing, and facultatively anaerobic Actinobacterium. The growth temperature ranges from 15 to 40°C (optimal at 37°C) and the growth pH range from 6 to 9 (optimal at 8). It was isolated from a 297 m depth-drilling core obtained from the Iberian Pyrite Belt (Puente-Sánchez et al., 2014a). Only the innermost part of the core was sampled, and sodium bromide was added to the drilling water as a tracer for potential contamination, as described in Amils et al. (2013).

Two members of the *Tessaracoccus* genus have been previously isolated from marine sediments (Lee and Lee, 2008) and deep subsurface environments (Finster et al., 2009). Others have been isolated from crude oil-contaminated saline soil (Cai et al., 2011) and oleaginous, water-mixed metalworking fluids (Kämpfer et al., 2009), which suggests that the *Tessaracoccus* genus might be specially adept at degrading hydrocarbons and/or recalcitrant organic matter under harsh environmental conditions.

Culture Medium

The composition of the anoxic medium FE used in this study was (wt/vol): 0.25% NaCl; 0.04% NH_4Cl; 0.003% $MgCl_2 \cdot 6H_2O$; 0.005% $CaCl_2 \cdot 2H_2O$; 0.2% $FeCl_2 \cdot 4H_20$; 0.01% yeast extract; 0.085% $NaNO_3$; 0.1% glucose; 0.1% succinic anhidryde; 0.05% KH_2PO_4; 0.025% $NaHCO_3$; 0.05% cysteine hydrochloride; 0.01% resazurin. The pH of the medium was 6 and it was sterilized at 121°C for 20 min.

Study of Crystal Nucleation and Precipitation

T. lapidicaptus was inoculated into liquid cultures which were carried out in 100 ml bottles containing 100 ml of FE medium. The bottles containing the culture medium were incubated anaerobically at 30°C and examined periodically for the presence of minerals for up to 45 days after incubation. Controls consisting of uninoculated culture media and media inoculated with non-viable cells were included in all the experiments. pH measurements were performed at the end of the growth and mineral formation.

The optical density (OD) of the inocula was 0.5 at a wavelength of 600 nm. It was analyzed using a Spectronic 20 Genesys spectrophotometer. The Fe^{2+} was measured using a RQflex 10 Merck reflectoquant.

Mineral Analysis

The crystals were examined by X-ray diffraction (XRD) using a PANlytical X'Pert MPD PW3011/10. A JEOL JSM 6335 scanning electron microscope (SEM), equipped with a spectroscope of dispersive energy (EDX), was used for imaging and elemental

analysis of single crystals. The mineral precipitates were also analyzed by transmission electro microscopy (TEM) under JEOL JEM 2100, 200 KV TEM with a CCD camera model 832. The morphology of the cells and crystal precipitates were examined with a JEOL JEM-1010 (TEM). TEM sample preparation is described in Ferrero et al. (2013) but without adding 5% $K_3Fe(CN)_6$ in the post-fixation step.

Geochemical Studies

The activity of dissolved species and the degree of saturation in the solutions assayed were determined using the geochemical computer program PHREEQC version 2 (Parkhust and Appelo, 1999). The results from PHREEQC are presented in terms of the saturation index (SI) for each predicted mineral. SI is defined by SI = lg (IAP/Ksp), where IAP is the ion activity product of the dissolved mineral constituents in a solubility product (Ksp) for the mineral. Thus, SI > 0 implies oversaturation with respect to the mineral, whereas SI < 0 means undersaturation.

Results

The mineral precipitates formed exclusively in culture bottles with active bacterial cells, while no mineral precipitation occurred in sterile parallel controls (bottles with non-viable cells and without cells). The pH changed from 6 to ~7.5 in cultures with living bacteria. No change in pH was detected in the control experiments. The starting concentration of Fe^{2+} was 0.56 g/L and the final concentration, after mineral precipitation, 0.01 g/L. The XRD study reveals that the bioprecipitates are composed of vivianite and siderite, being vivianite the dominant mineral phase (**Figure 1**).

TEM and SEM images of the bacterial precipitates show that Fe-phosphate crystals and Fe-carbonate spheroidal nanoparticles (nanoglobules) and in some cases, elongated nanoparticles were attached to the bacterial cells and EPS (**Figures 2A–D**, **3A,B,D**, **4A,B**). EDX analyses (**Figures 2F–H**) confirm the X-ray results, the nanoparticle precipitates are composed of both, vivianite and siderite. Vivianite crystals have a prismatic or tabular habit and form coarse radial-fibrous aggregates like rosettes with a high degree of crystallinity and vitreous luster (**Figures 3C,D**). These crystals are approximately 10–20 μm in width and 100–300 μm in length. Siderite crystals are aggregates of nanoglobules with a diameter 20–100 nm (**Figures 4A,B**). These nanoglobules were attached to *T. lapidicaptus* cells and embedded in a thin organic film (exopolymeric substances or EPS) produced by *T. lapidicaptus* during its growth (**Figures 2A–D**, **4A,B**). Mineralized bacteria were clearly recognized (**Figures 3A**, **4A**) as well as dividing cells (**Figures 2B**, **4A**); broken cells and mould of degraded cells (**Figures 3A**, **4B**). The process of microspherulites (diameter > 10 μm) formation comprises a sequence of events, starting with the appearance of bacterial nanoglobules (<20 nm) to larger ones (>100 nm), which agglomerate with time resulting in microspherulites (**Figures 4C,D**). The most important process in the sequence that leads to the formation of spherulites is the accumulation of nanoglubules and mineralized bacterial cells, embedded in EPS matrix, displaying a granulated texture (**Figures 4A,B,D**).

Mineral phases with SI values positive or very close to 0 (above or below the equilibrium point) were observed, suggesting the possibility for inorganic (chemical) precipitation in the aqueous medium assayed (**Table 1**). These SI data were obtained by applying the geochemical software PHREEQC to the ionic composition of the culture medium. According to these data, the culture FE medium is saturated in hydroxiapatite, vivianite, and siderite.

Discussion

Nucleation and Precipitation of Phosphate and Carbonate by *T. Lapidicaptus*

Our TEM and SEM studies showed that carbonate and phosphate nanocrystals nucleated on bacterial cell surfaces and EPS (**Figures 2**, **3**, **4**). The initial step of nucleation of carbonate and phosphate spheroidal nanocrystals occurs in the outer side of the bacterial envelopes (cell wall) and within EPS in intimate association with the bacteria cell surface (**Figures 2A–D**, **3A**, **4A,B**). Later carbonate spherulites and elongated phosphate crystals are formed by aggregation of nanocrystals embedded in the EPS matrix (**Figures 2E**, **3A,B**, **4A,B,D**). We also observed mineralized bacterial cells embedded in the surface of the crystals (**Figures 3A**, **4A**). Similar nanocrystals have been previously reported for culture experiments using sulfate reducing-bacteria, aerobic heterotrophic bacteria, acidophilic iron-reducing bacteria and fungi (Aloisi et al., 2006; Bontognali et al., 2008; Sánchez-Román et al., 2008, 2014; Oggerin et al., 2013). These findings lead us to propose that microbial nanostructures such as nanocrystals and crystalline nanoparticles are not related to a single microbial group or to a specific microbial metabolism but to a wide range of microorganisms including bacteria and fungi. However, the mineralogy composition of such nanostructures would depend on the physico-chemical properties of the precipitating solution (chemistry, pH, salinity, etc.,) and on the type of microorganism involved in the precipitation.

The aqueous culture medium used in these experiments is saturated with respect to vivianite and siderite (**Table 1**), there is a tendency toward their abiotic precipitation. The calculated saturation indexes (**Table 1**) for vivianite and siderite indicate that they should have been abiotically precipitated in the solution, in the absence of bacteria. However, no mineral precipitation was observed in the control experiments (without and with non-viable cells), while in living culture experiments (with active cells) vivianite and siderite precipitated. These data confirm that an aqueous solution saturated with certain mineral phase(s) does not imply abiotic precipitation with respect to those minerals, but it depends on their precipitation kinetics (Morse, 1983). Therefore, vivianite and siderite precipitation can be attributed to the presence of living *T. lapidicaptus* cells which are capable of overcoming the kinetic barriers for mineral precipitation, i.e., reducing the activation energy barriers. The metabolic activity of the bacteria is very important because it supplies the ions necessary for the formation of minerals, PO_4^{3-} for phosphates and CO_3^{2-} for carbonates. Additionally, the appropriate microenvironment around bacterial cells and EPS (increase pH

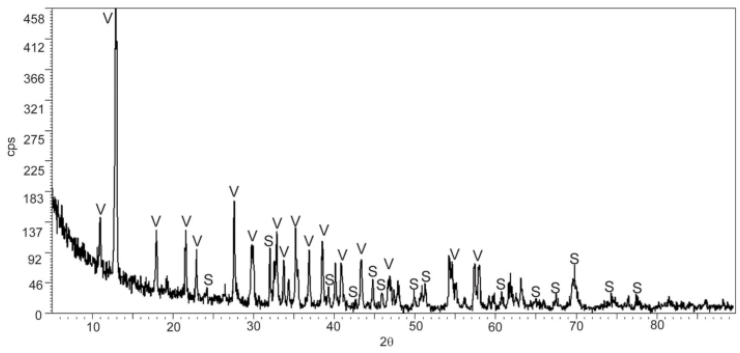


FIGURE 1 | X-ray diffractogram of the bioprecipitates formed in *T. Lapidicaptus* anaerobic cultures. V, Vivianite; S, siderite.

and/or ionic concentration) is created for mineral precipitation. In fact, an increase in the pH from 6 to 7.8 in the cultures with active cells was measured. No mineralization was observed in the control experiments where no pH alteration was detected.

Bacteria induce mineral precipitation by concentrating ions (e.g., Ca, Fe, Mg, CO_3^{2-}, PO_3^{4-}, NH_4^+) and changing the pH in the microenvironment surrounding their cells (Ehrlich, 2002; van Lith et al., 2003; Sánchez-Román et al., 2011). Bacterial cells act as a template for mineral nucleation by adsorbing ions around the cellular surface membrane or cell wall (Schultze-Lam et al., 1996; Bosak and Newman, 2003). This process does not occur in absence of bacterial activity. Moreover, EPS are considered as important factor for mineral precipitation (Dupraz et al., 2004; Aloisi et al., 2006; Ercole et al., 2007; Bontognali et al., 2008; Krause et al., 2012). The charged cell walls as well as the reactive groups of the EPS provide active interfacial sites for adsorption and complexation of dissolved aqueous metal species, inducing the nucleation and precipitation of minerals by reducing the activation energy barriers (Konhauser, 1998; De Yoreo et al., 2013; Habraken et al., 2013). This results in a mineralized cellular matrix containing detectable concentrations of metallic ions that are not easily re-dissolved (Beveridge and Fyfe, 1985). During bacterial growth experiments using an aqueous medium rich in organic compounds (yeast extract, glucose, succinic anhidryd, cysteine) source for CO_2, NH_3 and HPO_3^{2-}, the pH, carbonate and phosphate concentrations increased because of the production of CO_2, NH_3, and HPO_3^{2-} (which hydrate to form CO_3^{2-}, NH_4^+ and PO_4^{3-}) during metabolization of organic compounds.

These changes together with the adsorption of iron ions by *T. lapidicaptus* would lead to local supersaturation gradients around bacterial surfaces and EPS, using these as nucleation sites to induce iron phosphate and carbonate precipitation as previously has been demonstrated for other type of bacteria (Aloisi et al., 2006; Bontognali et al., 2008; Sánchez-Román et al., 2008, 2011, 2014).

It is essential to understand how super-saturation and nucleation develop in our culture experiments. If concentration of ions in solution exceeds the solubility product for a solid mineral phase, precipitation will not occur until a certain degree of supersaturation is achieved (Berner, 1980). The process during which the maximum free energy is attained is known as nucleation and involves the growth of critical crystal nuclei (Jack et al., 1993; Sánchez-Navas et al., 2009). This process is accompanied by a decrease in free energy and is referred to as crystal growth. The presence of living bacteria can promote either process. Here, the surface of the microbial cell and EPS provides a template on which nucleation can occur and overcome kinetic barriers to facilitate mineral precipitation as previously expounded. This reduces the free energy required during the nucleation step and focuses crystal growth because nucleation on the biological template occurs before nucleation in homogeneous solution (Jack et al., 1993). Most types of bacteria are capable of acting as nucleation templates (Beveridge and Fyfe, 1985). Although abiotic precipitation is difficult in natural systems or in sterile laboratory experiments (present work), the presence of bacteria can induce the precipitation of minerals in microenvironments by (1) modifying the conditions of their surrounding environments and/or concentrate ions in the bacterial cell envelope and (2) acting as nucleation sites.

Apparently, siderite and vivianite have many physico-chemical characteristics in common. However, morphological details suggest different nucleation and growth conditions. In most cases, siderite nucleates on the bacterial cell wall and with time develops into a microspherulite (**Figures 2A–C**, **4A–D**); and vivianite nucleates on the bacterial cell and within the EPS on nanocrystals that agglomerate and get alone as elongated (rosette) crystals (**Figures 2A–E**, **3A–D**). On the other hand, PO_4^{3-} inhibits the precipitation of carbonate minerals (Bouropoulos and Koutsoukos, 2000; Kofina and Kotsoukos, 2005; Morse et al., 2007). Whereas, the presence of phosphate can inhibit the formation of siderite (Fredrickson and Gorby, 1996). Indeed, when phosphate is present, vivianite appears to be the stable end product due to its lower solubility product (Ksp $= 10^{-36}$) (Glasauer et al., 2003). In our cultures, vivianite precipitated first than siderite. Organic

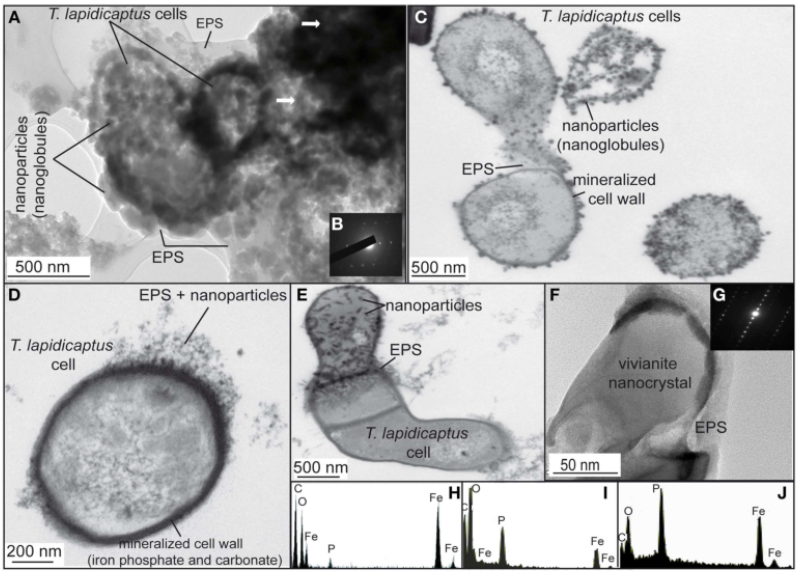


FIGURE 2 | TEM images of the bioprecipitates formed in *T. Lapidicaptus* anaerobic cultures. (A) The bioprecipitates are nanoparticles attached to *T. lapidicaptus* cell and its secreted EPS, respectively. **(B)** Electron diffraction pattern of the darker areas, more mineralized areas. **(C)** *T. lapidicaptus* cells with mineralized cell wall and covered by nanoparticles embeded in EPS. **(D)** Detail of *T. lapidicaptus* cell with mineralized cell wall. Note nanoparticles embedded in EPS. **(E)** Detail of three cells together with mineralized cell wall and EPS. The upper cell covered by nanoparticles embedded in EPS. **(F)** Elongated nanoparticle, vivianite nanocrystal, embedded in EPS. **(G)** Electron diffraction pattern of the nanocrystal **(F)**. **(H,I)** EDX spectra of both dark and lighter mineralized areas (1A) composed of Fe-carbonate and phosphate, respectively. **(J)** EDX spectrum of nanoparticle (1E) composed of Fe-phosphate (vivianite).

phosphates are hydrolysed by phosphatases, which liberate orthophosphate during microbial decomposition of organic material. Locally elevated orthophosphate, excreted during microbial decomposition of organic material (yeast extract), around bacterial cells becomes available together with the inorganic phosphate and soluble iron (Fe^{2+}) initially present in the culture medium for the precipitation of vivianite. Hence, the precipitation of Fe-phosphate (vivianite) removes PO_4^{3-} ions from the solution, leading to the precipitation of Fe-carbonate (siderite). Thus, we propose that vivianite and siderite are authigenic sedimentary minerals that require similar physico-chemical conditions to precipitate and the presence of microorganisms.

Significance and Implications of Vivianite and Siderite in Natural Systems

In our aqueous solutions, conditions for the precipitation of Fe-carbonates are created after phosphate precipitation. Thus, we propose that in environments with sufficient phosphate and iron, vivianite will precipitate first than siderite, while in environments with PO_4^{3-} deficiency siderite will precipitate first. The same phenomenon might be occurring in Rio Tinto subsurface, from where *T. lapidicaptus* has been isolated (Puente-Sánchez et al., 2014a), characterized by the presence of iron sulfide (pyrite) and carbonate (siderite) minerals (Fernández-Remolar et al., 2012). However, Fe-phosphates have not yet been detected there, even though vivianite is considered as an alteration of pre-existing Fe-carbonates or sulfides (Garvin, 1998). This could be due to the presence of abundant dissolve sulfide, which inhibits the formation of vivianite (Postma, 1981; Manning et al., 1999). The reduction of aqueous or embedded sulfate coupled to organic matter oxidation would lead to the formation of H_2S and carbonate. H_2S subsequently reacts with iron, and contributes to the formation of FeS_2 whereas CO_3^{2-} would react with iron to precipitate carbonate ($FeCO_3$). Then, the absence of phosphate minerals in Rio Tinto may be linked to the presence of sulfate (Puente-Sánchez et al., 2014b) and its microbiological transformation to H_2S with the consequently formation of pyrite. It is probably for this reason that we rarely find vivianite occurring in nature. In fact, it only persists in reducing organic-rich environments (lakes, deep-sea sediments, swamps, sewage) with low concentrations of sulfate, which results in a high and continuous precipitation of vivianite. Our experimental findings provide information that could be used to interpret the role of microorganisms in digenetic mineral processes resulting in phosphate and carbonate formation in natural systems.

On the other hand, supersaturated solutions (e.g., interstitial pore water) cannot serve as reliable predictors for the *in situ* formation of phosphates (Rothe et al., 2014). A gel-like pore structure (Rothe et al., 2014) of a sediment matrix rich in organic

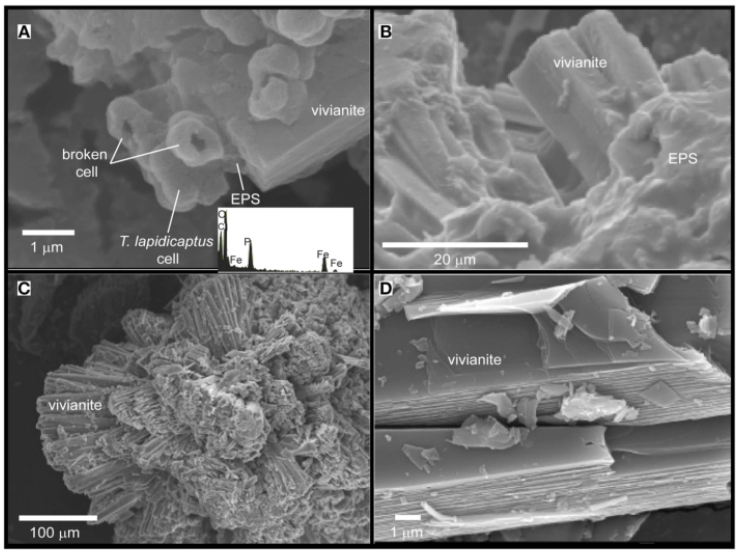


FIGURE 3 | SEM images of the Fe-phosphate precipitates from *T. Lapidicaptus* anaerobic cultures. EDX spectrum of mineralized cell displaying C, O, Fe, and P. **(A)** Vivianite crystal attached to mineralized *T. lapidicaptus* cells and EPS. **(B)** Elongated vivianite crystal embedded in EPS. **(C)** Rosette formation of crystal clusters of vivianite. **(D)** Vivianite crystals with prismatic or tabular habit.

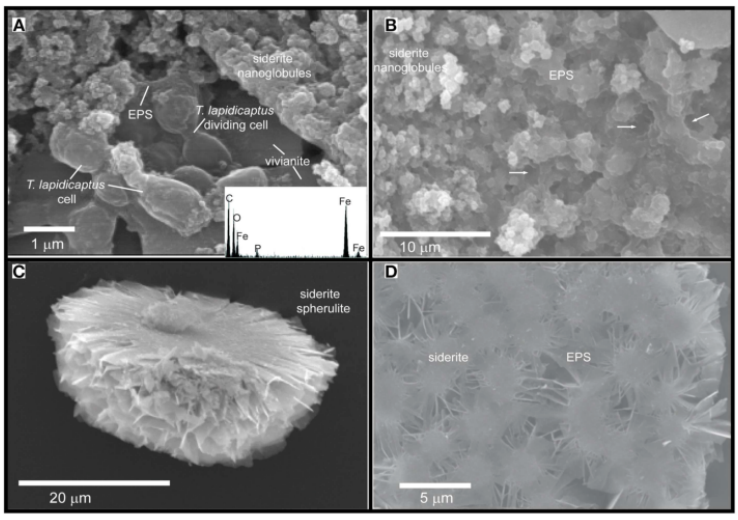


FIGURE 4 | SEM images of the Fe-carbonate precipitates from *T. Lapidicaptus* anaerobic cultures. (A) Siderite nanoglobules embedded in EPS and attached to mineralized dividing *T. Lapidicaptus* cells. Note the vivianite crystal attached to these cells. EDX spectrum of mineralized cell displaying C, O, Fe, and small peak of P. **(B)** Fe-carbonate nanoglobules (siderite) embedded in EPS and delimiting the bacterial cell contours (white arrows). These nanostructures display granulated texture. White arrows correspond to moulds of degraded bacteria (broken cells). **(C)** Broken microspherulite of siderite. **(D)** Detail of a siderite spherulite which formed by aggregation of nanoparticles.

TABLE 1 | Saturation index values (SI) in FE anaerobic medium.

Mineral phase	SI
Aragonite, $CaCO_3$	−1.16
Artinite, $Mg_2(CO_3)(OH)_2 \times 3H_2O$	−8.45
Brucite, $Mg(OH)_2$	−6.76
Calcite, $CaCO_3$	−1.02
Dolomite, $CaMg(CO_3)_2$	−1.34
Halite, NaCl	−4.61
Huntite, $CaMg_3(CO_3)_4$	−5.98
Hydroxiapatite $Ca_5(PO_4)_3OH$	2.48
Magnesite, $MgCO_3$	−0.81
Natron, $Na_2CO_3 \times 10H_2O$	−7.46
Nesquehonite, $MgCO_3 \times 3H_2O$	−3.22
Siderite, $FeCO_3$	2.40
Vivianite, $Fe_3(PO_4)_2 \times 8H_2O$	9.08

Results are from geochemical software PHREEQC. SI values are for initial conditions.

matter, in combination with release of soluble phosphorous and iron due to microbial activity, is necessary for vivianite formation in natural systems. In our culture experiments the gel-like pore structure would be the bacterial EPS and cell surfaces which throughout bacterial activity (1) create local microenvironments supersaturated with respect to phosphates and carbonates; and (2) act as templates for mineral nucleation overcoming the kinetic barriers of mineral precipitation. Indeed, similar nanocrystals of phosphate and carbonate to the ones shown in the present study (**Figures 2A,E**, **4A,B**) have been also reported as closely linked to the presence of bacterial cells, EPS and similar mucilaginous structures in modern and ancient environments (Sánchez-Román et al., 2008; Crosby and Bailey, 2012; Cosmidis et al., 2013, 2014; Sánchez-Navas et al., 2013). The preservation of these nanostructures in the geological record may help us to trace microbial processes through geologic time. Therefore, this experimental study (1) may help to understand the formation of ancient iron phosphate (Fife and Mark, 1982; Cook and Shergold, 1986; Simonen, 1986) and carbonate (Veizer et al., 1989; Ohmoto et al., 2004) deposits, and (2) provides potential biosignatures that may be useful to test terrestrial and extraterrestrial habitats for life evidences.

Finally, our experiments demonstrate that *T. lapidicaptus* can cause considerable iron accumulation through biomineralization of phosphate and carbonate, therefore, *T. lapidicaptus* could be considered a good phosphate removing bacterium from anaerobic systems. Furthermore, vivianite and siderite produced by *T. lapidicaptus* could be a good alternative fertilizer of phosphorous and iron. The presence of this bacterium and/or related bacteria in natural environments could explain the formation of vivianite and siderite. The co-precipitation of Fe-phosphate and carbonate in our cultures links the P, C, and Fe cycles during biomineralization.

Author Contributions

MS and FP designed the culture experiments and performed all the laboratory tasks, carried out the culture experiments. MS wrote the first draft of the manuscript. FP, VP, and RA assisted in preparing the manuscript, all authors read and approved the final version.

Acknowledgments

This work was supported by the European Research Council Advanced Grant ERC-250350/IPBSL, and the Spanish Ministry of Economy and Competitiveness AYA2011-24803 projects. We also acknowledge Nuria Rodriguez for her assistance with some of the SEM analyses.

References

Aloisi, G., Gloter, A., Krúger, M., Wallmann, K., Guyot, F., and Zuddas, P. (2006). Nucleation of calcium carbonate on bacterial nanoglobules. *Geology* 34, 1017–1020. doi: 10.1130/G22986A.1

Amils, R., Fernández-Remolar, D., Parro, V., Rodríguez-Manfredi, J. A., Timmis, K., Oggerin, M., et al. (2013). Iberian Pyrite Belt Subsurface Life (IPBSL), a drilling project of biohydrometallurgical interest. *Adv. Mat. Res.* 825, 15–18. doi: 10.4028/www.scientific.net/AMR.825.15

Berner, R. A. (1980). *Early Diagenesis*. Princeton, NJ: Princeton University Press.

Beveridge, T. J., and Fyfe, W. S. (1985). Metal fixation by bacterial cell walls. *Can. J. Earth Sci.* 22, 1892–1898. doi: 10.1139/e85-204

Bontognali, T. R. R., Vasconcelos, C., Warthmann, R., Dupraz, C., Bernasconi, S. M., and McKenzie, J. A. (2008). Microbes produce nanobacteria-like structures, avoiding cell entombment. *Geology* 36, 663–666. doi: 10.1130/G24755A.1

Bouropoulos, N. C., and Koutsoukos, P. G. (2000). Spontaneous precipitation of struvite from aqueous solutions. *J. Crys. Growth* 213, 381–388. doi: 10.1016/S0022-0248(00)00351-1

Bosak, T., and Newman, D. K. (2003). Microbial nucleation of calcium carbonate in the Precambrian. *Geology* 31, 577–580. doi: 10.1130/0091-7613(2003)031%3C0577:MNOCCI%3E2.0.CO;2

Cai, M., Wang, L., Cai, H., Li, Y., Wang, Y. N., Tang, Y. Q., et al. (2011). Salinarimonas ramus sp. nov. and *Tessaracoccus oleiagri* sp. nov., isolated from a crude oil-contaminated saline soil. *Int. J. Syst. Evol. Microbiol.* 61, 1767–1775. doi: 10.1099/ijs.0.025932-0

Cook, P. J., and Shergold, J. H. (1986). *Phosphate Deposits of the World: Vol. 1: Proterozoic and Cambrian Phosphorites*. Cambridge: Cambridge University Press.

Cosmidis, J., Benzerara, K., Gheerbrant, E., Estève, I., Bouya, B., and Amaghzaz, M. (2013). Nanometer-scale characterization of exceptionally preserved bacterial fossils in paleocene phosphorites from ouled abdoun (Morocco). *Geobiology* 11, 139–153. doi: 10.1111/gbi.12022

Cosmidis, J., Benzerara, K., Morin, G., Busigny, V., Lebeau, O., Jézéquel, D., et al. (2014). Biomineralization of iron-phosphates in the water column of Lake Pavin (Massif Central, France). *Geochim. Cosmochim. Acta* 126, 78–96. doi: 10.1016/j.gca.2013.10.037

Craig, J. R., and Vaughan, D. J. (1994). *Ore Microscopy and Ore Petrography, 2nd Edn*. New York, NY: Wiley.

Crosby, C. H., and Bailey, J. V. (2012). The role of microbes in the formation of modern and ancient phosphatic mineral deposits. *Front. Microbiol.* 3:241. doi: 10.3389/fmicb.2012.00241

De Yoreo, J. J., Waychunas, G. A., Jun, Y. S., and Fernandez-Martinez, A. (2013). *In situ* investigations of carbonate nucleation on mineral and organic surfaces. *Rev. Mineral. Geochem.* 77, 229–257. doi: 10.2138/rmg.2013.77.7

Dijkstra, N., Kraal, P., Kuypers, M. M. M., Schnetger, B., and Slomp, C. P. (2014). Are iron-phosphate minerals a sink for phosphorous in anoxic black sea sediments? *PLoS ONE* 9:e101139. doi: 10.1371/journal.pone.0101139

Dupraz, C., Reid, R. P., Braissant, O., Decho, A. W., Norman, R. S., and Visscher, P. T. (2009). Processes of carbonate precipitation in modern microbial mats. *Earth Sci. Rev.* 96, 141–162. doi: 10.1016/j.earscirev.2008.10.005

Dupraz, C., Visscher, P. T., Baumgartner, L. K., and Reid, R. P. (2004). Microbe–mineral interactions: early carbonate precipitation in a hypersaline lake (Eleuthera Island, Bahamas). *Sedimentology* 51, 745–765. doi: 10.1111/j.1365-3091.2004.00649.x

Dyar, M. D., Jawin, E. R., Breves, E., Marchand, G., Nelms, M., Lane, M. D., et al. (2014). Mössbauer parameters of iron in phosphate minerals: implications for interpretation of martian data. *Am. Mineral.* 99, 914–942, doi: 10.2138/am.2014.4701

Ehrlich, H. L. (2002). *Geomicrobiology, 4th Edn.* New York, NY: Marcel Dekker.

Ellwood, B. B., Chrzanowski, T. H., Hrouda, F., Long, G. J., and Buhl, M. L. (1998). Siderite formation in anoxic deep-sea sediments: a synergetic bacteria controlled process with important implications in paleomagnetism. *Geology* 16, 980–982.

Ercole, C., Cacchio, P., Botta, A. L., Centi, V., and Lepidi, A. (2007). Bacterially induced mineralization of calcium carbonate: the role of exopolysaccharides and capsular polysaccharides. *Microsc. Microanal.* 13, 42–50. doi: 10.1017/S1431927607070122

Eynard, A., del Campillo, M. C., Barron, V., and Torrent, J. (1992). Use of vivianite ($Fe_3(PO_4)_2 \times 8H_2O$) to prevent iron chlorosis in calcareous soils. *Fert. Res.* 31, 61–67. doi: 10.1007/BF01064228

Fairen, A., Fernández-Remolar, D., Dohm, J. M., Baker, V. R., and Amils, R. (2004). Inhibition of carbonates synthesis in acidic oceans on early Mars. *Nature* 431, 423–426. doi: 10.1038/nature02911

Fernández-Remolar, D. C., Preston, L. J., Sánchez-Román, M., Izawa, M. R. M., and Huang, L. (2012). Carbonate precipitation under bulk acidic conditions as a potential biosignature for searching life on Mars. *Earth. Planet. Sci. Lett.* 351–352, 13–26. doi: 10.1016/j.epsl.2012.07.015

Ferrero, J. J., Alvarez, A. M., Ramírez-Franco, J., Godino, M. C., and Bartolomé-Martín, D. (2013). β-Adrenergic receptors activate exchange protein directly activated by cAMP (Epac), translocate Munc13-1, and enhance the Rab3A-RIM1a interaction to potentiate glutamate release at cerebrocortical nerve terminals. *J. Biol. Chem.* 288, 31370–31385. doi: 10.1074/jbc.M113.463877

Fife, D. L., and Mark, E. B. (1982). Significance of Neogene phosphorites in Capistrano embayment, Southern California. *AAPG Bull.* 66, 1688.

Finster, K. W., Cockell, C. S., Voytek, M. A., Gronstal, A. L., and Kjeldsen, K. U. (2009). Description of *Tessaracoccus profundi* sp.nov., a deep-subsurface actinobacterium isolated from a Chesapeake impact crater drill core (940 m depth). *Ant. van Leeuwen.* 96, 515–526. doi: 10.1007/s10482-009-9367-y

Fredrickson, J. K., and Gorby, Y. A. (1996). Environmental processes mediated by iron-reducing bacteria. *Curr. Opin. Biotechnol.* 7, 287–294. doi: 10.1016/S0958-1669(96)80032-2

Fredrickson, J. K., Zachara, J. M., Kennedy, D. W., Dong, H., Onslott, T. C., Hinman, N. W., et al. (1998). Biogenic iron mineralization accom- panying the dissimilatory reduction of hydrous ferric oxide by a groundwater bacterium. *Geochim. Cosmochim. Acta* 62, 3239–3257. doi: 10.1016/S0016-7037(98)00243-9

Garvin, P. (1998). *Iowa's Minerals: Their Occurrence, Origins, Industries, and Lore.* Iowa City, IA: University of Iowa Press.

Glasauer, S., Weidler, P. G., Langley, S., and Beveridge, T. J. (2003). Controls on Fe reduction and mineral formation by a subsurface bacterium. *Geochim. Cosmochim. Acta* 67, 1277–1288. doi: 10.1016/S0016-7037(02)01199-7

Habraken, W. J., Tao, J., Brylka, L. J., Friedrich, H., Bertinetti, L., Schenk, A. S., et al. (2013). Ion-association complexes unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nat. Commun.* 4, 1507. doi: 10.1038/ncomms2490

Jack, T. R., Ferris, F. G., Stehmeier, L. G., Kantzas, A., and Marentette, D. F. (1993). "Bug rock: bacteriogenic mineral precipitation systems for oil patch use," in *Developments in Petroleum Science in Microbial Enhancement of Oil Recovery – Recent Advances,* eds E. T. Premuzic and A. Woodhead (Amsterdam, NY: Elsevier), 27–35.

Jakobsen, B. H. (1988). Accumulation of pyrite and Fe-rich carbonate and phosphate minerals in a lowland moor area. *J. Soil Sci.* 39, 447–455. doi: 10.1111/j.1365-2389.1988.tb01230.x

Kämpfer, P., Lodders, N., Warfolomeow, I., and Busse, H. J. (2009). *Tessaracoccus lubricantis* sp. nov., isolated from a metalworking fluid. *Int. J. Syst. Evol. Microbiol.* 59, 1545–1549. doi: 10.1099/ijs.0.006841-0

Kholodov, V. N., and Butuzova, G. Y. (2008). Siderite formation and evolutions of sedimentary iron ore deposition in the Earth's history. *Geol. Ore Deposits* 50, 299–319. doi: 10.1134/S107570150804003X

Kofina, A. N., and Kotsoukos, P. G. (2005). Spontaneous precipitation of struvite from synthetic wastewater solutions. *Cryst. Growth Des.* 5, 489–496. doi: 10.1021/cg049803e

Konhauser, K. O. (1998). Diversity of bacteria iron mineralization. *Earth Sci. Rev.* 43, 91–121. doi: 10.1016/S0012-8252(97)00036-6

Krause, S., Liebetrau, V., Goeb, S., Sánchez-Román, M., McKenzie, J. A., and Treude, T. (2012). Microbial nucleation of Mg-rich dolomite in exopolymeric substances under anoxic modern seawater salinity: new insight into an old enigma. *Geology* 40, 587–590. doi: 10.1130/G32923.1

Last, W. M., and De Deckker, P. (1990). Modern and Holocene carbonate sedimentology of two saline volcanic maar lakes, southern Australia. *Sedimentology* 37, 967–981. doi: 10.1111/j.1365-3091.1990.tb01839.x

Lee, D. W., and Lee, S. D. (2008). Tessaracoccus flavescens sp. nov., isolated from marine sediment. *Int. J. Syst. Evol. Microbiol.* 58, 785–789. doi: 10.1099/ijs.0.64868-0

Lee, S. Y., Yoshikawa, H., and Matsui, T. (2010). Biomineralization of vivianite on the carbon steel surface attacked by the iron reducing bacteria. *MRS Proc.* 1265:AA06-01. doi: 10.1557/proc-1265-aa06-01

Lemos, V. P., Lima da Costa, M., and Lemos, R. L. (2007). Vivianite and siderite in lateritic iron crust: an example of bioreduction. *Quim. Nova* 30, 36–40. doi: 10.1590/S0100-40422007000100008

Lovley, D. R., Phillips, E. J. P., and Lonergan, D. J. (1991). Enzymatic versus nonenzymic mechanisms for iron (III) reduction in aquatic sediments. *Environ. Sci. Technol.* 25, 1062–1067 doi: 10.1021/es00018a007

Manning, P. G., Murphy, T. P., and Prepas, E. E. (1991). Intensive formation of vivianite in the bottom sediments of mesotrophic Narrow Lake, Alberta. *Can. Mineral.* 29, 77–78.

Manning, P. G., Prepas, E. E., and Serediak, M. S. (1999). Pyrite and vivianite intervals in the bottom sediments of eutrophic Baptiste lake, Alberta, Canada. *Can. Mineral.* 37, 593–601.

McGowan, G., and Prangnell, J. (2006). The significance of vivianite in archeological settings. *Geoarcheology* 21, 93–111. doi: 10.1002/gea.20090

Mikhailov, V. (1940). Vivianite as fertilizer for industrial potatoes. *Spirto-Vodoch. Prom.* 17, 14–15.

Moore, C. H. (1989). "The nature of carbonate depositional systems-comparison of carbonates and siliciclastics," in *Developments in Sedimentology* 46, ed C. H. Moore (Amsterdam: Elsevier), 1–19.

Morse, J. W. (1983). "The kinetics of calcium carbonate dissolution and precipitation," in *Reviews in Mineralogy: Carbonates-Mineralogy and Chemistry*, ed R. J. Reeder (Chelse, MA: Mineralogical Society of America, Bookcrafters, Inc.), 227–264.

Morse, J. W., Arvidson, R. S., and Lüttge, A. (2007). Calcium carbonate formation and dissolution. *Chem. Rev.* 107, 342–381. doi: 10.1021/cr050358j

Mortimer, R. J. G., and Coleman, M. (1997). Microbial influence on the oxygen isotopic composition of diagenetic siderite. *Geochim. Cosmochim. Acta* 61, 1705–1711. doi: 10.1016/S0016-7037(97)00027-6

Nelipa, K. V. (1961). Vivianite - a source of phosphorous for plant nutrition. *Uchenye Zapiski Michurink Gosndarst Pedagog Inst.* 8, 65–80.

Nriagu, J. O., and Dell, C. I. (1974). Diagenetic formation of iron phosphates in recent lake sediments. *Am. Mineral.* 59, 934–946.

Nutt, M. L., and Swihart, G. H. (2012). Phosphate mineralization at Nonconnah Creek, Menphis, Tennessee. *J. Tennessee Acad. Sci.* 87, 126.

Oggerin, M., Tornos, F., Rodríguez, N., del Moral, C., Sánchez-Román, M., and Amils, R. (2013). Specific jarosite biomineralization by *Purpureocillium lilacinum*, an acidophilic fungi isolated from Río Tinto. *Environ. Microbiol.* 15, 2228–2237. doi: 10.1111/1462-2920.12094

Ohmoto, H., Watanabe, Y., and Kumazawa, K. (2004). Evidence from massive siderite beds for a CO2-rich atmosphere before ~ 1.8 billion years ago. *Nature* 429, 395–399. doi: 10.1038/nature02573

Orange, F., Westall, F., Disnar, J. R., Prieur, D., Bienvenu, N., Le Romancer, M., et al. (2009). Experimental silicification of the extremophilic Archaea Pyrococcus abyssi and *Methanocaldococcus jannaschii*: applications in the

search for evidence of life in early Earth and extraterrestrial rocks. *Geobiology* 7, 403–418. doi: 10.1111/j.1472-4669.2009.00212.x

Parkhust, D. L., and Appelo, C. A. J. (1999). *User's Guide to Phreeqc (version 2) – a Computer Program for Speciation, Batchreaction, One-Dimensional Transport, and Inverse Geochemical Calculations*. Water-Resources Investigations Report, 99–4259, US Geological Survey, Denver, CO.

Postma, D. (1981). Formation of siderite and vivianite and the porewater composition of a recent bog sediment in Denmark. *Chem. Geol.* 31, 225–244. doi: 10.1016/0009-2541(80)90088-1

Puente-Sánchez, F., Moreno-Paz, M., Rivas, L. A., Cruz-Gil, P., García-Villadangos, M., Gómez, M. J., et al. (2014b). Deep subsurface sulfate reduction and methanogenesis in the Iberian Pyrite Belt revealed through geochemistry and molecular biomarkers. *Geobiology* 12, 34–47. doi: 10.1111/gbi.12065

Puente-Sánchez, F., Sánchez-Román, M., Amils, R., and Parro, V. (2014a). Tessaracoccus lapidicaptus sp. *nov*., a novel actinobacterium isolated from the deep subsurface of the Iberian Pyritic Belt (Huelva, Spain). *Int. J. Syst. Evol. Microbiol.* 64, 3546–3552. doi: 10.1099/ijs.0.060038-0

Raiswell, R., and Canfield, D. (2012). The iron biogeochemical cycle past and present. *Geochem. Perspect.* 1, 1–220. doi: 10.7185/geochempersp.1.1

Rakshit, S., Matocha, J. C., and Coyne, M. S. (2008). Nitrite reduction by siderite. *Soil Sci. Soc. Am. J.* 72, 1070–1077. doi: 10.2136/sssaj2007.0296

Ridgwell, A., and Zeebe, R. E. (2005). The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth Planet. Sci. Lett.* 234, 299–315. doi: 10.1016/j.epsl.2005.03.006

Riding, R. (2006). Microbial carbonate abundance compared with fluctuations in metazoan diversity over geological time. *Sediment. Geol.* 185, 229–238. doi: 10.1016/j.sedgeo.2005.12.015

Rothe, M., Frederichs, T., Eder, M., Kleeberg, A., and Hupfer, M. (2014). Evidence for vivianite formation and its contribution to long-term phosphorus retention in a recent lake sediment: a novel analytical approach. *Biogeosciences* 11, 5169–5180. doi: 10.5194/bg-11-5169-2014

Sánchez-Alcalá, I., del Campillo, M. C., Barrón, V., and Torrent, J. (2012). Post evaluation of synthetic nanosiderite for the prevention of iron chlorosis. *J. Sci. Food Agric.* 92, 1964–1973. doi: 10.1002/jsfa.5569

Sánchez-Navas, A., Martín-Algarra, A., Rivadeneyra, M. A., Melchor, S., and Martín Ramos, J. D. (2009). Crystal-growth behavior in Ca-Mg carbonate bacterial spherulites. *Cryst. Growth Des.* 9, 2690–2699. doi: 10.1021/cg801320p

Sánchez-Navas, A., Martín-Algarra, A., Sánchez-Román, M., Jiménez-López, C., Nieto, F., and Ruiz-Bustos, A. (2013). "Crystal growth of inorganic and biomediated carbonates and phosphates," in *Advanced Topics on Crystal Growth*, ed S. Olavo Ferreira (Rijeka: InTech), 67–88.

Sánchez-Román, M., Fernández-Remolar, D., Amils, R., Sánchez-Navas, A., Schmid, T., San Martin-Uriz, P., et al. (2014). Microbial mediated formation of Fe-carbonate minerals under extreme acidic conditions. *Sci. Rep.* 4:4767. doi: 10.1038/srep04767

Sánchez-Román, M., Romanek, C. S., Fernández-Remolar, D. C., Sánchez-Navas, A., McKenzie, J. A., Amils, R., et al. (2011). Aerobic biomineralizacion of Mg-rich carbonates: implications for natural environments. *Chem. Geol.* 281, 143–150. doi: 10.1016/j.chemgeo.2010.11.020

Sánchez-Román, M., Vasconcelos, C., Schmid, T., Dittrich, M., McKenzie, J. A., Zenobi, R., et al. (2008). Aerobic microbial dolomite at the nanometer scale: implicatios for the geologic record. *Geology* 36, 879–882. doi: 10.1130/G25013A.1

Sapota, T., Aldahan, A., and Al-Asam, I. S. (2006). Sedimentary facies and climate control on formation of vivianite and siderite microconcretions in sediments of Lake Baikal, Siberia. *J. Paleolimmnol.* 36, 245–257. doi: 10.1007/s10933-006-9005-x

Schultze-Lam, S., Fortin, D., Davis, B. S., and Beveridge, T. J. (1996). Mineralization of bacterial surfaces. *Chem. Geol.* 132, 171–181 doi: 10.1016/S0009-2541(96)00053-8

Simonen, A. (1986). Vivianite from Paakkila, Tuusniemi, Finland. *Bull. Geol. Soc. Finl.* 58, 271–275.

Thali, M. J., Lux, B., Lösch, S., Rösing, F. W., Hürlimann, J., Feer, P., et al. (2011). "Brienzi" - The blue Vivianite man of Switzerland: time since death estimation of an adipocere body. *Forensic Sci. Int.* 211, 34–40. doi: 10.1016/j.forsciint.2011.04.009

Tomkinson, T., Lee, M. R., Mark, D. F., and Smith, C. L. (2013). Sequestration of Martian CO_2 by mineral carbonation. *Nature Comm.* 4, 2662. doi: 10.1038/ncomms3662

Valley, J. W., Eiler, J. M., Graham, C. M., Gibson, E. K. Jr., Romanek, C. S., and Stolper, E. M. (1997). Low-temperature carbonate concretions in the martian meteorites ALH 84001: evidence from stable isotopes and mineralogy. *Science* 275, 1633–1638.

van Lith, Y., Warthmann, R., Vasconcelos, C., and McKenzie, J. A. (2003). Sulphate-reducing bacteria induce low-temperature dolomite and high Mg-calcite formation. *Geobiology* 1, 71–79. doi: 10.1046/j.1472-4669.2003.00003.x

Veeramani, H., Alessi, D. S., Suvorova, E. I., Lezama-Pacheco, J. S., Stubbs, J. E., Sharp, J. O., et al. (2011). Products of abiotic U(VI) reduction by biogenic magnetite and vivianite. *Geochim. Cosmochim. Acta* 75, 2512–2528. doi: 10.1016/j.gca.2011.02.024

Veizer, J., Hoefs, J., Lowe, D. R., and Thurston, P. C. (1989). Geochemistry of precambrian carbonates: II. Archean greenstone belts and Archean sea water. *Geochim. Cosmochim. Acta* 53, 859–871. doi: 10.1016/0016-7037(89)90031-8

Vuillemin, A., Ariztegui, D., De Coninck, A. S., Lücke, A., Mayr, C., Schubert, C. J., et al. (2013). Origin and significance of diagenetic conditions in sediments of Laguna Potrok Aike, southern Argentina. *J. Paleolimnol.* 50, 275–291. doi: 10.1007/s10933-013-9723-9

Wiberg, E., Wiberg, N., and Holleman, A. F. (2001). *Inorganic Chemistry*. San Diego, CA: Academic Press.

Wilkinson, M., Haszeldine, R. S., Fallick, A. E., and Osborne, M. J. (2000). Siderite zonation within the Brent Group: microbial influence or aquifer flow? *Clay Miner.* 35, 111–121. doi: 10.1180/000985500546512

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies

Nucleic Acids Research Advance Access published November 8, 2015

Nucleic Acids Research, 2015 **1**
doi: 10.1093/nar/gkv1113

A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies

Fernando Puente-Sánchez[1,*], Jacobo Aguirre[1,2,3] and Víctor Parro[1]

[1]Department of Molecular Evolution, Centro de Astrobiología (INTA-CSIC). Instituto Nacional de Técnica Aeroespacial, Ctra de Torrejón a Ajalvir km 4. 28850 Torrejón de Ardoz, Madrid, Spain, [2]Centro Nacional de Biotecnología (CSIC). c/ Darwin 3, 28049 Madrid, Spain and [3]Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain

Received April 20, 2015; Revised September 08, 2015; Accepted October 12, 2015

ABSTRACT

Adequate read filtering is critical when processing high-throughput data in marker-gene-based studies. Sequencing errors can cause the mis-clustering of otherwise similar reads, artificially increasing the number of retrieved Operational Taxonomic Units (OTUs) and therefore leading to the overestimation of microbial diversity. Sequencing errors will also result in OTUs that are not accurate reconstructions of the original biological sequences. Herein we present the Poisson binomial filtering algorithm (PBF), which minimizes both problems by calculating the error-probability distribution of a sequence from its quality scores. In order to validate our method, we quality-filtered 37 publicly available datasets obtained by sequencing mock and environmental microbial communities with the Roche 454, Illumina MiSeq and IonTorrent PGM platforms, and compared our results to those obtained with previous approaches such as the ones included in mothur, QIIME and USEARCH. Our algorithm retained substantially more reads than its predecessors, while resulting in fewer and more accurate OTUs. This improved sensitiveness produced more faithful representations, both quantitatively and qualitatively, of the true microbial diversity present in the studied samples. Furthermore, the method introduced in this work is computationally inexpensive and can be readily applied in conjunction with any existent analysis pipeline.

INTRODUCTION

High-throughput sequencing of marker genes, such as the 16S ribosomal RNA, has become an invaluable tool for microbial ecologists, since it allows for a previously unreachable level of detail in the analysis of complex microbial communities. Many studies have used platforms such as the Roche 454, Illumina or IonTorrent sequencers to thoroughly characterize and compare microbial communities at an affordable cost (1–5), while others have taken advantage of their very high yield in order to analyze the structure and composition of the rare biosphere (6). However, the correct assessment of sequencing artifacts is critical in obtaining representative results. Reads derived from the same biological template may differ due to sequencing errors, which can cause them to be assigned to different clusters and therefore result in the overestimation of microbial diversity (7). The most common software tools and packages include sequence clustering into OTUs in their recommended pipelines (8–15, see (15) for a comparison of several molecular ecology pipelines). Alternatives to traditional clustering have been recently proposed, such as distribution-based clustering (16) or a clustering-free approach (17). These novel methods are specially suited for subpopulation level studies, but work only for moderate-to-high abundance sequences, being unsuitable for population-level alpha or beta diversity studies (17). Moreover, even although they can remove likely erroneous sequences and resolve subpopulations based on dynamic information, they nevertheless rely on a quality filtering step for the preprocessing of raw reads (17).

Amplicon denoising (18,19) is a widespread method for filtering Roche 454 pyrosequencing reads that can also be applied to IonTorrent data. It works on flowgrams rather than sequences, which allows for a more natural modeling of the homopolymer read errors that are characteristic of pyrosequencing and ion semiconductor sequencing. However, it is platform specific and computationally expensive.

For Illumina systems, there is no consensus approach to quality filtering, with the authors of mothur (20), QIIME (21) and UPARSE (15) proposing different solutions. All those heuristic approaches were published as parts of their respective pipelines, but to the best of our knowledge they have not been thoroughly compared to each other.

The lack of a rigorous method for incorporating quality scores in the analysis of marker-gene sequences has also led some authors to advocate for a stringent filtering in or-

*To whom correspondence should be addressed. Email: fpusan@gmail.com

der to reduce the retrieval of spurious diversity (3). However, overstringent filtration will result in an undesired loss of sensitivity and will have an impact on the observed taxonomic distribution (21). Therefore, an accurate algorithm that overcomes these problems is desirable.

Herein we present and validate the *Poisson binomial filtering* (PBF) method, which is able to determine the error-probability distribution of any sequence with associated quality scores, by using a simple statistical approach. Phred quality scores, which represent the probability that a given base call is mistaken, can be derived from the raw output of every sequencing platform. Reading a single base can be likened to tossing a coin: the base is either right or wrong, and both chances can be determined from its quality score. In fact, the number of errors present in a given base follows a Bernoulli distribution, i.e. a binomial distribution with a single trial. For a sequence of nucleotides with potentially non-equal error probabilities, we sum their associated Bernoulli random variables in order to obtain the estimated probability that the sequence has accumulated more than k errors, where k is the maximum number of errors that still allows for a correct clustering (Supplementary Note 1). When compared with the filtering approaches included in mainstream molecular ecology pipelines such as mothur, QIIME or UPARSE, Poisson binomial filtering proved to be the most accurate algorithm for filtering marker-gene sequences. Additionally, PBF is based on simple statistical principles and, since it only requires Phred quality scores as an input, it is expected to work robustly regardless of the sequencing platform, assuming that the quality scores are acceptable predictors of the true error probabilities, and that errors are independent. Finally, our algorithm is computationally efficient, scales linearly with the number of sequences, and has a low memory fingerprint, making it useful even in low-performance desktop environments.

MATERIALS AND METHODS

The Poisson binomial filtering algorithm

Let us suppose we have 1 sequence of length N nucleotides (nt), each nucleotide with a potentially non-equal probability p_i of being erroneous and a probability (1-p_i) of being correct. Our target is to obtain the probability of this sequence of having j erroneous nucleotides, for $j = 0, 1, 2, \ldots, N$ (see example in Figure 1a,b). Statistically, our problem can be analyzed as the probability distribution of the number of successes in a sequence of N independent yes/no experiments with success probabilities $p_1, p_2, \ldots, p_N$. This is equivalent to the sum S_N of N independent Bernoulli distributed random variables $X_1, X_2, \ldots, X_N$ such that $S_N = \sum_{i=1}^{N} X_i$, where

$$\begin{aligned} P(X_i = j) &= 1 - p_i \text{ for } j = 0, \\ P(X_i = j) &= p_i \quad \text{for } j = 1, \\ P(X_i = j) &= 0 \quad \text{for } j > 1, \end{aligned} \tag{1}$$

and $P(X_i = j)$ stands for the probability of obtaining j errors in nucleotide i. The stochastic variable S_N follows a Poisson binomial distribution (Supplementary Note 1), from where we name the method presented here.

While the probability of obtaining a sequence with j errors in a sequence, for all values of j, can be expressed explicitly (see Eq. (SN1.2) and its derivation in section Supplementary Note SN1.1), it becomes useless in practice for moderate values of j. We explain here an alternative algorithm inspired by (22) that allows us to calculate the error-probability distribution $P(S_N = j)$ for all j in a simple and efficient way.

First, note that if we have two random variables Y and Z, each of them taking discrete values 0, 1, 2,…, the probability of the sum $Y+Z$ of taking value j is

$$P(Y + Z = j) = \sum_{i=0}^{j} P(Y = i)\,P(Z = j - i). \tag{2}$$

The algorithm results:

1. Obtain $P(X_1 = j)$ from Equation (1). Let $U = X_1$.
2. For $i = 2, 3, \ldots, N$, the distribution is obtained by following (a–c) recursively.
   (a) Calculate $P(X_i = j)$ from Equation (1).
   (b) Calculate $P(Y+Z = j)$ from Equation (2), being $Y = U$ and $Z = X_i$.
   (c) Let $U = Y+Z$.
3. The estimated probability for the sequence under study of having j errors, $P(S_N = j)$, is given by U when $i = N$.
4. The steps (1–3) must be repeated for $j = 0, 1, 2, \ldots, j_{\max}$, where $j_{\max}$ is the lowest value of j that satisfies $\sum_{r=0}^{j} P(S_N = r) \geq \xi$ and $0 < \xi < 1$ is a confidence coefficient (in our case $\xi = 0.995$). Let j_ξ be the number such that the sequence has a probability ξ of having less than j_ξ errors. It is obtained interpolating the accumulated error probability of the sequence between the values $r = j_{\max}$-1 and $r = j_{\max}$ to obtain its exact value in $r = j_\xi$. A linear interpolation yields
$$j_\xi = j_{max} - 1 + \frac{\xi - \sum_{r=0}^{j_{\max}-1} P(S_N = r)}{P(S_N = j_{max})}.$$
5. Let j_{tol} be the maximum tolerable number of errors per sequence, that is, the maximum number of errors allowed for a correct clustering. In our calculations, we have fixed j_{tol} at 1% of the trimmed sequence length (that is, at 2.5 for the 454 and Illumina datasets, and at 2 for the Ion-Torrent datasets). The sequence under study is discarded if $j_\xi > j_{\text{tol}}$, and accepted otherwise (Figure 1b,c). At this moment, the calculation for this particular sequence is finished, and it is time to repeat the whole algorithm for the rest of the sequences of the population.

j_ξ is therefore the predicted maximum number of errors of a given sequence, with a confidence coefficient of ξ. Under the parameters used in this study, a sequence has an *alpha* $= (1 - \xi) = 0.005$ probability of having more than j_ξ errors (assuming that Q scores accurately reflect true error probabilities, see discussion for more details). In practice, this also means that j_ξ will be an overestimation of the actual number of errors. This can be easily avoided by the users by setting a different confidence coefficient. However, we also want to

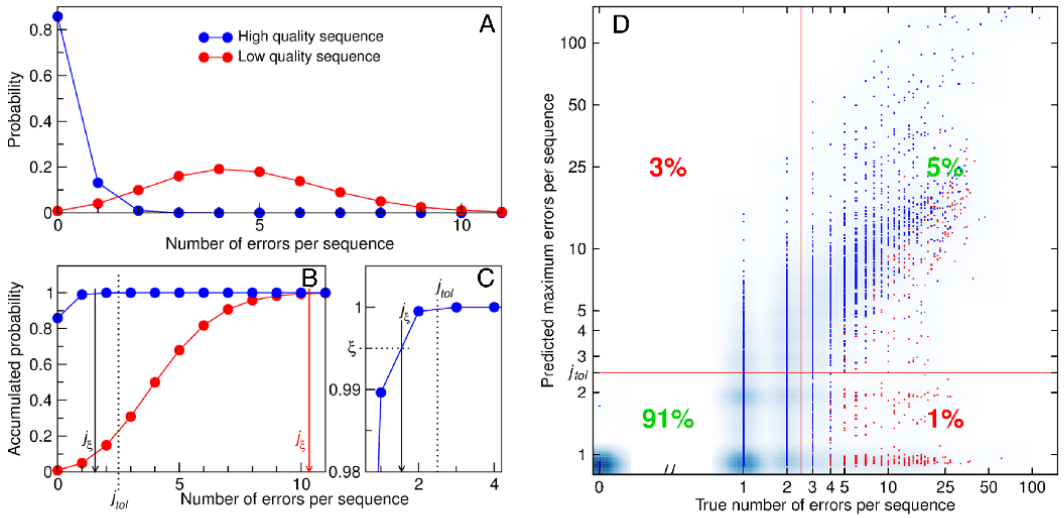


Figure 1. Poisson binomial filtering accurately discriminates between good and erroneous sequences. (**A, B**): Error probability distribution (**A**) and accumulated error probability distribution (**B**) of two example nucleotide sequences, as calculated from their quality scores by the Poisson binomial filtering algorithm. j_ξ stands for the 99.5th percentile of the error probability (i.e. a sequence has a probability $\xi = 0.995$ of having less than j_ξ errors). j_{tol} is the maximum tolerable number of errors (1% of the sequence length in our case). Sequences with $j_\xi > j_{tol}$ are discarded in the filtering step. (**C**) Zoom of (**B**), sketching the calculation of j_ξ for the high quality sequence. (**D**) Comparison between the predicted maximum errors j_ξ calculated by the Poisson binomial algorithm and the true number of errors for all sequences from the Even1T mock community dataset. Dots represent unique sequences. True mock community sequences are plotted in blue, contaminant sequences are plotted in gray, and chimeric sequences are plotted in red. The blue background represents sequence abundance (note that few unique sequences may have a high number of representatives, and vice versa). Red lines indicate our error cut-off of 2.5 errors per sequence (j_{tol}). The plot is thus divided in four quadrants corresponding to correctly retained sequences (lower left), correctly discarded sequences (upper right), incorrectly discarded sequences (upper left) and incorrectly retained sequences (lower right). The percentage of true mock community sequences present on each quadrant is also indicated. PBF correctly classified 96% of the non-chimeric/non-contaminant sequences present in the Even1T dataset. The graph is plotted in logarithmic scale (the 0 in the x-axis is added for clarity).

note that our recommended parameters provide a clear upper bound to the number of errors which is unlikely to be trespassed, and nonetheless result in a very high proportion of correctly classified sequences (see results and Figure 1d).

Finally, as our problem corresponds to the sum of N binomial distributions of probabilities p_i and number of trials $n = 1$, it can be approximated to a Poisson distribution as far as N is high and $p_i << 1$. The Poisson approximated probability for the sequence under study of having j errors, $P(S_N = j)$, becomes

$$P(S_N = j) = \frac{\lambda^j \exp(-\lambda)}{j!},$$

where

$$\lambda = \sum_{i=1}^{N} p_i .$$

While this approximation is reasonably accurate and quicker to compute, it may fail even for high-quality sequences, provided they contain one or more low-quality bases. This occurrence is not uncommon in real datasets, and can lead to significant differences between the Poisson binomial filtering algorithm and its Poisson approximation (Supplementary Note SN1.5).

A more detailed explanation of the Poisson binomial filtering algorithm presented above and its Poisson approximation can be found in Supplementary Note 1.

Algorithm implementation

Both C and Python implementations of the Poisson binomial filtering algorithm are available in GitHub (http://github.com/fpusan/moira), and as a pip-installable python package (which can be installed by typing 'pip install moira' in a linux, mac or windows command line). The execution time of the Poisson binomial algorithm and its Poisson approximation for increasing numbers of sequences and for sequences of increasing length can be found in Supplementary Note 8. Full documentation is available at https://github.com/fpusan/moira/blob/master/README.md. The Poisson binomial filtering algorithm has also been included in the LotuS OTU processing pipeline (23, http://psbweb05.psb.ugent.be/lotus/).

The moira filtering pipeline

The script *moira.py* contains an implementation of the Poisson binomial filtering algorithm and performs the following tasks:

- If required, it assembles contigs from paired reads (*–paired*). The assembler is an implementation of mothur make.contigs command (http://www.mothur.org/wiki/Make.contigs), and includes a modified version of the Needleman-Wunsch global aligner and a consensus sequence constructor. Our implementation also returns consensus quality scores, which are simply the highest quality scores for each position of the alignment.
- It truncates sequences to a fixed length (*–truncate*), discarding the sequences that are smaller than the cut-off.
- It applies the Poisson binomial algorithm to calculate the predicted maximum number of errors of each remaining sequence, with a given confidence coefficient (settable by the *–alpha* parameter) and discards the ones that have more errors per nucleotide than the specified cut-off (*–uncert*). The *alpha* parameter is defined as $1 - \xi$, and represents the probability of underestimating the errors present on a given sequence.
- It collapses identical sequences and chooses the one with the least predicted maximum errors as the group representative for filtering (*–collapse*). We assumed that, in spite of differences in quality, identical sequences should have the same origin, as it is unlikely that two biologically unrelated sequences become identical due to sequencing errors. Thus, if one of them has good quality, the rest should be considered as true biological sequences and be allowed into the final dataset. We have demonstrated that collapsing sequences prior to quality filtering actually helps to mitigate an important source of taxonomic bias during sequence processing (Supplementary Note 2).

16S mock community data

Two synthetic mock microbial communities designed by the Human Microbiome Project (24, http://www.hmpdacc.org/HMMC) were used for evaluating the different filtering methods. Genetic DNA from 22 different organisms (20 bacterial, 1 archaeal and 1 eukaryotic) was mixed in known amounts, based on qPCR of the small subunit (SSU) rRNA gene, in order to generate two different mixtures: an Even mock community, in which there is a similar amount of SSU rRNA copies for each organism, and a Staggered mock community, in which the amounts of SSU rRNA of each organism are different.

The data used in this study come from publicly available libraries generated by sequencing the Even and Staggered mock communities with the Roche 454 GS FLX Titanium, the Illumina MiSeq and the IonTorrent PGM platforms. References for all the datasets used in this study are given in Supplementary Note 3.

Validation of Poisson binomial filtering on mock community data

The script *moira.py* was used to predict the number of errors present on each sequence for all the six Roche 454 GS FLX Titanium, the four Illumina MiSeq and the two IonTorrent PGM mock community datasets. For the MiSeq datasets, contigs were first assembled from paired-end reads by applying the *–paired* flag. The *–alpha* parameter, which indicates the probability of a read having more errors than reported, was left as its default value of 0.005. Identical reads were collapsed and the sequence with the smallest number of errors was chosen as the group representative, as described above. These predicted maximum values were compared to the true number of errors of each sequence, which was obtained by using the mothur command *seq.error*. Briefly, the sequences were aligned to a reference database made up from the true biological sequences present in the mock community (which can be found in http://www.mothur.org/wiki/454_SOP). Sequences with less than 80% alignment coverage were discarded at this step. The resulting alignment was then used to determine the true number of errors present on each sequence, as well as whether that sequence was chimeric or not, bearing in mind that chimeras cannot be identified with absolute certainty. Likely non-chimeric sequences that nevertheless showed less than 95% similarity to their best hit in the mock reference database were aligned again against mothur's SILVA 16S reference alignment (25, version 98). In case said sequence showed a pairwise identity and an alignment coverage equal or greater to 95% to any sequence in the 16S reference alignment, it was considered to be a contaminant.

Quality filtering of 16S reads

USEARCH. Trimming of reads by quality values was performed by using the USEARCH *fastq_filter* command, as employed by (15). Reads (for 454/IonTorrent data) or contigs (for paired Illumina data) were truncated at the first position with a quality score below 15 (*–fastq_trunqual* 15). After that, sequences were truncated to a length of 250 nucleotides (200 nucleotides for IonTorrent data), and sequences smaller than 250 nt (200 nt for IonTorrent data) were discarded (*–fastq_trunclen* 250/200).

We also tested a different method implemented in the USEARCH *fastq_filter* command, as suggested in the author's web page (http://drive5.com/usearch/manual/uparse_cmds.html). Briefly, reads (for 454 data and IonTorrent) or contigs (for paired MiSeq data) with more than 0.5 expected errors (*–fastq_maxee* 0.5) were discarded. After that, sequences were truncated to a length of 250 nt (200 nt for IonTorrent data), and sequences smaller than 250 nt (200nt for IonTorrent data) were discarded (*–fastq_trunclen* 250/200). NOTE: During the review process of this manuscript, more details on the USEARCH expected errors method were published in (26). We now discuss the differences and similarities between the USEARCH expected errors filter and our own in Supplementary Note 9.

Mothur. Denoising of 454 and IonTorrent reads was performed using the mothur command *shhh.flows*, which is an implementation of the PyroNoise algorithm, as recommended in the mothur SOP (www.mothur.org/wiki/Analysis_examples) After denoising, mothur command *trim.seqs* was used to truncate the denoised sequences to a length of 250 nt, and to discard sequences smaller than 250 nt (200 nt in both cases for IonTorrent data).

Additionally, paired Illumina reads were assembled and filtered according to mothur's MiSeq SOP (http://www.

mothur.org/wiki/MiSeq_SOP). After filtering, mothur command *trim.seqs* was used to truncate the contigs to a length of 250 nt, and to discard sequences smaller than 250 nt.

QIIME. Paired Illumina reads were assembled with QIIME's *join_paired_ends.py* script, using the default parameters. QIIME's script *split_libraries_fastq.py* was used to filter the resulting contigs, as recommended by the authors (*-r* 3 *-p* 0.75 *-q* 3 *-n* 0) in (21).

Poisson binomial filtering. The script *moira.py* was used to perform Poisson binomial filtering on 454/IonTorrent reads or contigs assembled from Illumina paired reads (*–paired*), as described above. Before filtering, sequences or contigs were truncated to 250 nt, and the sequences smaller than 250 nt (200 nt in both cases for IonTorrent reads) were discarded (*–truncate* 250/200). Identical 454/IonTorrent reads or Illumina contigs were clustered together prior to quality control (*–collapse*) and the sequence with the highest quality was chosen as the group representative for quality control. 0.01 or less errors per nucleotide were tolerated (*–uncert* 0.01) with a 0.005 chance of error underestimation (*–alpha* 0.005).

For each method, paired Illumina reads were assembled as recommended by its authors.

Note that, for consistency, we have chosen the 250 nt cutoff recommended by (15) as the fixed length for the rest of the filtering methods, for the 454 and Illumina datasets. Since read length may have an effect in clustering and OTU accuracy, we believe that equalizing it results in more valid comparisons between the different filtering methods. In a similar fashion, the 200 nt cutoff proposed in http://www.brmicrobiome.org/#!16sprofilingpipeline/cuhd was applied to all the filtering methods for the IonTorrent datasets.

The full list of commands used for each method can be found in Supplementary Note 10.

Common processing pipeline for the filtered reads

Regardless of the filtering method, the filtered sequences were subjected to a common pipeline based in mothur's recommended SOP that included the following steps:

- Sequence alignment to mothur's SILVA Reference Alignment.
- Optimization of the alignment space by removing the sequences that failed to align correctly, in order to ensure that all the remaining sequences overlap at the same region of the SILVA Reference Alignment.
- Pre-clustering of similar sequences.
- Removal of chimeras with UCHIME.
- Taxonomic classification and removal of non-bacterial and unclassified sequences.
- Library size standardization (see below).
- Clustering of the remaining sequences using mothur's default average neighbor algorithm, with an OTU distance cut-off of 0.03.
- Accuracy classification of the resulting OTUs (see below).

For each sample, the libraries obtained after filtering the raw reads with the different methods were standardized to a similar size by random sub-sampling. Total number of retrieved OTUs and singleton OTUs, as well as accuracy assessment of the OTU representative sequences, were obtained by averaging the results from 100 independent rounds of random library size standardizations followed by clustering of the resulting reads.

The full list of commands can be found in Supplementary Note 10.

OTU accuracy assessment on mock communities

The accuracy of the obtained OTU representative sequences was evaluated by aligning them to a reference database made up from the true biological sequences present in the sample, as previously described by (15). Sequence alignment was performed with mothur *align.seqs* command. If the pairwise identity of an OTU representative sequence to any sequence in the reference database was 100%, the OTU was classified as 'Perfect'. If the pairwise identity was smaller than 100%, but greater or equal to 99%, the OTU was classified as 'Good'. If the pairwise identity was smaller than 99% but greater or equal to 97%, the OTU was classified as 'Noisy'. If the pairwise identity was lower than 95%, the OTU representative sequence was aligned to mothur's SILVA bacterial 16S reference alignment (version 98). If said sequence showed a pairwise identity and an alignment coverage equal or greater to 95% to any sequence in the 16S reference alignment, the OTU was classified as 'Contaminant'. When none of the above conditions applied, the OTU was considered to be the result of either an undetected chimera, a mock community sequence with more than 3% errors or a novel contaminant sequence with more than 5% errors, and was classified as 'Other'.

OTU accuracy assessment on environmental communities

OTU representative sequences from environmental communities were aligned with mothur's SILVA bacterial 16S reference alignment (25). For each dataset and filtering method, the average similitude of the OTU representative sequences to their best hits in the SILVA alignment was calculated. This was taken as an indicator of the overall accuracy of the resulting OTUs, under the assumption that sequencing errors are more likely to decrease OTU similitude to known sequences than to increase it.

References for all the environmental datasets used in this study are given in Supplementary Note 3.

Assessment of the taxonomic bias caused by the different filtering methods

Taxonomic bias was assessed by comparing the taxonomic composition of the sample before and after performing quality filtering. Sequences were classified by using the *classify.seqs* command implemented in mothur and mothur's RDP 16S rRNA reference database (version 9). Then, taxonomic composition was obtained by calculating the proportion of sequences that were assigned to each phylotype at the genus level with an 80% confidence cut-off (40% for the environmental communities). Finally, taxonomic bias was calculated as the Bray-Curtis dissimilarity between the filtered and unfiltered sequence communities. In the 454 and

IonTorrent libraries from the environmental communities, a high proportion of sequences did not get classified at the genus level. Therefore, the taxonomic composition of those libraries was instead calculated at the class level.

RESULTS

We validated the Poisson binomial filtering algorithm and compared it with the different filtering approaches recommended by the authors of mothur (8,13,20), USEARCH-UPARSE (10,15,26) and QIIME (21) by quality-filtering datasets obtained by sequencing different mock and environmental microbial communities with the Roche 454 GS FLX Titanium, the Illumina MiSeq and the IonTorrent PGM platforms. In order to evaluate the different methods on equal grounds, filtered reads were processed with a common downstream pipeline that included chimera-filtering with UCHIME (27), sample size standardization and OTU clustering.

PBF accurately discriminates between good and erroneous sequences

When applying our default cut-off of 1% errors allowed per sequence, our algorithm accurately classified 96% of the mock community sequences from the Even1M dataset (Figure 1d). Three percent of the sequences were incorrectly discarded while, remarkably, only 1% of the sequences were incorrectly retained. Moreover, most of those incorrectly retained sequences had only three true errors (1.2% errors per sequence), meaning that they would likely cluster correctly when applying the standard 3% OTU distance cut-off. The rest of the Illumina datasets rendered similar results. The accuracy of our method was slightly lower for the 454 and IonTorrent datasets, but it nevertheless resulted in a minimum of 88% (for 454) and 79% (for IonTorrent) correctly classified sequences (Supplementary Figure SN4.1).

Performance of the different filtering methods on mock community datasets

Publicly available datasets from even and staggered mock communities from the Human Microbiome Project (24) were filtered with PBF, mothur, USEARCH and QIIME (Figure 2, Supplementary Note 4). These artificial communities contain known amounts of 16S rRNA gene copies from 20 different bacterial organisms. The fact that both the qualitative and quantitative composition of the samples are known beforehand allowed us to thoroughly compare the effects of the different filtering methods in terms of OTU accuracy, alpha diversity and community composition. OTU accuracy was defined as the maximum similarity of its representative sequence to the 16S sequences of the microorganisms used to build the mock community, as previously described in (15). We were also interested in determining how the different filtering processes affected the observed community composition. The taxonomic bias in community composition caused by any given filtering method was calculated as the Bray-Curtis dissimilarity between the raw and the filtered datasets, after taxonomically classifying their reads down to the genus level.

In the even datasets, which contain the same number of 16S rRNA gene copies for each organism, all methods resulted in more than 20 OTUs after clustering. This was not surprising, since contaminations, PCR errors and sequencing errors were expected to inflate the observed diversity. In the staggered communities, in which the number of 16S rRNA gene copies varied by several orders of magnitude between the different organisms, the observed diversity was generally lower, due to some species being present at very low abundances. The total number of reported OTUs greatly varied between filtering methods, with Poisson binomial filtering consistently resulting in values that were the closest to the true diversity of the samples.

PBF also produced the highest proportion of accurate OTUs in all the 16S mock datasets for both sequencing platforms, while minimizing the number of singletons and spurious OTUs retrieved (Figure 2a,b). In the 454 and IonTorrent datasets, it also discarded the smallest number of reads and resulted in the smallest taxonomic bias (Figure 2c,d). In the Illumina datasets QIIME retrieved a larger number of reads, while both QIIME and mothur caused smaller taxonomic biases than our method. (Figure 2c,d - Illumina). In the case of QIIME, we believe that this was the result of a too shallow filtering, since it produced a remarkably lower proportion of accurate OTUs and a larger number of OTUs and singletons (Figure 2a,b – Illumina). As for mothur, their method filters sequences based on the presence of mismatches of similar quality scores in aligned paired reads (20), which likely makes it less susceptible to biases in quality distribution between different taxonomic groups (Supplementary Note 2).

The two filtering algorithms included in the USEARCH suite showed an intermediate performance in terms of the number and accuracy of the OTUs retrieved for both the 454 and Illumina platforms. Quality trimming yielded the smallest number of reads and resulted in the highest taxonomic bias, which supports the idea that over-stringent filtering may lead to undesirable effects. In the IonTorrent datasets, USEARCH filtering performed below Poisson binomial filtering for all the studied benchmarks (Figure 2 - IonTorrent). Finally, the mothur implementation of the PyroNoise algorithm (12) showed lower OTU accuracy than the other methods tested for filtering 454 reads. It has been previously described that the denoising process can introduce minor alterations in the original reads (28), a phenomenon that might explain these results. It must be noted that, albeit a pipeline for filtering IonTorrent reads with PyroNoise has been described, the IonTorrent mock community datasets were only available in Fastq format (S. Salipante, personal communication), which precluded the use of flowgram denoising algorithms. However, this limitation was not present for the environmental datasets, and a comparison of quality filtering algorithms for IonTorrent datasets that includes PyroNoise can therefore be found in Supplementary Figure SN5.3.

It should be noted that both the QIIME and USEARCH-UPARSE pipelines include specific post-clustering steps that would have improved the results obtained by their filtering methods alone. QIIME recommends to apply a post-hoc OTU size cut-off to reduce the retrieval of spurious diversity (21) at the cost of sensitivity. USEARCH-UPARSE,

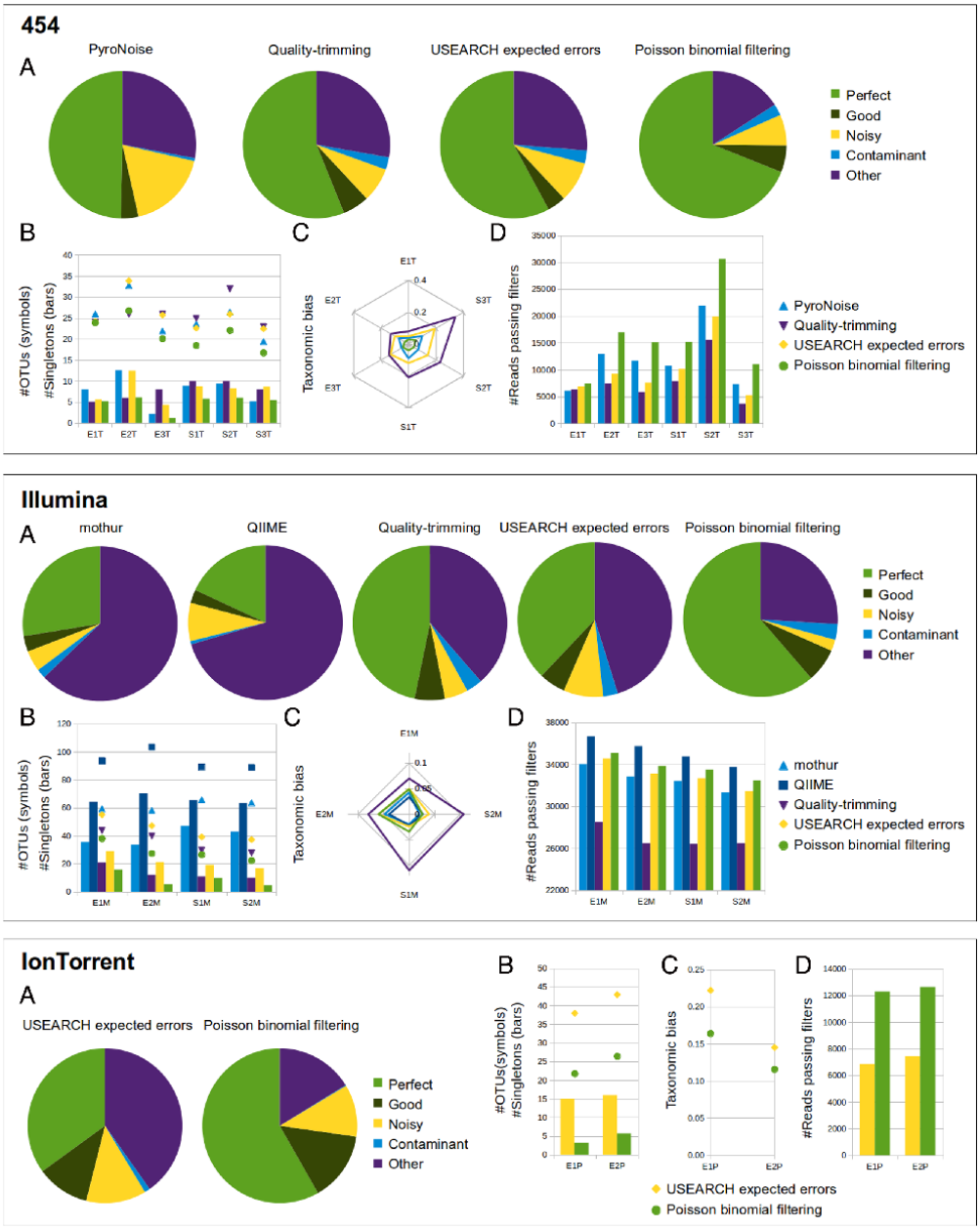


Figure 2. Comparison of filtering methods on 16S mock communities sequenced with the 454 GS FLX Titanium, Illumina MiSeq platforms and IonTorrent PGM platforms. (**A**) Pie charts constructed by averaging the fraction of OTUs on each accuracy category along the six 454 or the four Illumina samples. (**B**, **D**) Number of singletons (**B**, bars), total species (**B**, symbols) and reads (**D**) retrieved after filtering the raw reads with the different methods and performing chimera removal and clustering with a common pipeline. OTU and singleton numbers were obtained by averaging the results from 100 independent library size standardizations. (**C**) Taxonomic bias caused by the different filtering methods, measured as the Bray-Curtis dissimilarity between the raw and the filtered read communities.

in turn, maps the unfiltered reads to the OTU representative sequences after clustering in order to make an OTU table, and recovers reads with low predicted quality that are inferred to be good by comparison with high-quality sequences (15), increasing the number of retrieved reads. These steps are unrelated to quality-filtering *per se*, and can be applied independently of the method used to filter the sequences. Therefore, they were not included in our common post-filtering pipeline.

Performance of the different filtering methods on environmental datasets

The performance of the different filtering methods was also evaluated by quality-filtering publicly available datasets obtained by sequencing environmental communities (Supplementary Note 5). The results were similar to the ones obtained with the mock communities, with Poisson binomial filtering being the most consistent method in producing the smallest number of OTUs and singletons. Additionally, the OTUs obtained with PBF were overall the most similar to the 16S sequences present in the SILVA 16S reference alignment (25), which suggests that they contained the smallest number of errors. In the environmental 454 datasets, PyroNoise showed better results than in the 454 mock communities, but did it in an irregular fashion, especially in terms of OTU accuracy (Supplementary Figure SN5.1d). This inconsistency may be again due to the alteration of the original reads, and suggests that PyroNoise requires a finer parameter optimization than other approaches in order to be fully effective. In the environmental IonTorrent datasets PyroNoise discarded the smallest number of reads, but resulted in the highest number of singletons and OTUs, which also borne the least similarity to the reference alignment. USEARCH showed an intermediate performance between PyroNoise and Poisson binomial filtering (Supplementary Figure SN5.3). Finally, in the environmental Illumina datasets all filtering methods showed a similar behavior to that in the mock communities (Supplementary Figure SN5.2).

Quality-filtering is an additional source of taxonomic bias in microbial ecology studies

Even though the major sources of taxonomic biases in marker-gene-based studies are often related to differences in rRNA operon copy number, PCR and library construction (29–33), the read filtering process can increase this problem (Supplementary Note 2). We found significant biases in length and quality distribution between raw reads coming from different taxa in the mock 454 datasets (Figure 3a–c). Trimming them to a fixed length generated an artificial enrichment of the taxa with longer reads (Figure 3b), but since there is a decrease in quality at the end of 454 reads (see 15), it also resulted in a lower average read quality for the taxa with smaller raw reads (Figure 3d). This led to the generation of further taxonomic bias during the quality-filtering step (Figure 3f, Supplementary Note 2). Similar biases have been previously found in IonTorrent reads (34), and were confirmed during this study (Supplementary Note 2). Biases in read quality distribution between different taxa were also found for the mock Illumina datasets, although to a lesser extent. Such biases were remarkably higher than those due to the random removal of reads (Supplementary Note SN2.1, Supplementary Note 6). We solved this problem by collapsing identical reads and choosing the one with the highest quality as a representative for filtering, in order to decide whether the whole group was discarded or allowed into the filtered dataset. This procedure reduced the effect of quality distribution biases, as even low abundance sequences are expected to have a high quality representative. Our solution rendered similar quality distributions for the different taxa, even after length trimming (Figure 3e,f), and significantly lower taxonomic biases than other filtering approaches, especially for 454 data (Figure 2c). Every method that relies on quality scores for sequence filtering will be affected by this source of bias. We therefore propose the approach described above as a general solution to this problem, since its simplicity makes it very easy to integrate into any filtering pipeline.

DISCUSSION

In this work, we have presented and validated the Poisson binomial algorithm for filtering sequence reads based on their error probability distributions. We have also demonstrated that Poisson binomial filtering is especially useful in the context of gene-marker-based studies, such as the study of microbial populations by amplifying and sequencing their 16S rRNA gene.

We compared our algorithm with other five quality-filtering methods that are included as defaults in mainstream pipelines such as mothur, QIIME or USEARCH, by analyzing mock and environmental datasets generated with three different sequencing platforms. Our results show that, when coupled to a standard analysis pipeline that included chimera removal and clustering, PBF proved to be the most accurate algorithm for filtering marker-gene sequences. While retaining a large number of sequences, it also resulted in OTUs that were the closest to the true biological species present in the studied samples, and minimized the generation of spurious diversity and taxonomic biases. These metrics are useful for algorithm benchmarking, especially when coupled to the analysis of mock communities, but they are not the typical end result of molecular ecology studies. Instead, users are typically interested in features such as relative community richness, or patterns of community composition. In Supplementary Note 7 we show that the choice of filtering method can also have substantial effects on the retrieved relative community richness, potentially affecting the final ecological interpretation obtained from a given dataset.

Remarkably, this algorithm does not rely on any particular error model. Instead, it just derives the error probability distribution of a given sequence from the quality scores of its individual bases. The only assumptions that our algorithm makes are that for any given sequencing platform, sequencing errors will be independent and the quality scores obtained during base calling will truly represent the probabilities of that base being wrong. This conceptual simplicity is one of its main advantages: as long as accurate quality scores are provided, Poisson binomial filtering will work in

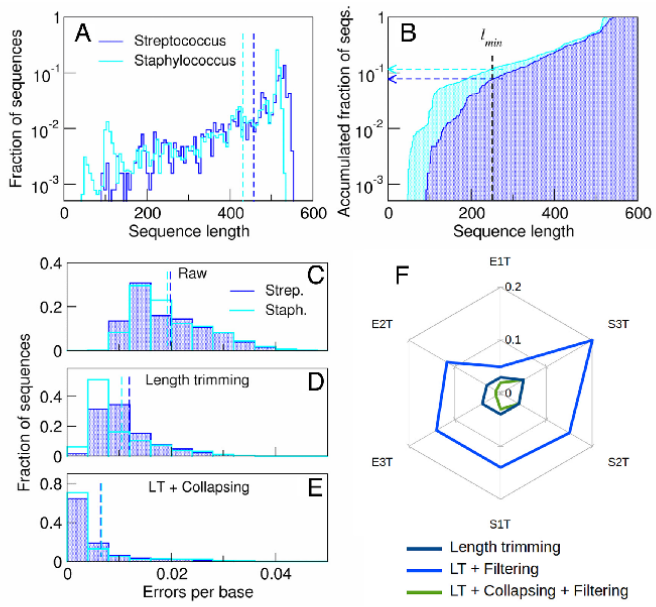


Figure 3. Addressing the taxonomic bias generated during the pre-processing and quality filtering of raw sequences. **(A, B)**: Raw reads from *Streptococcus* and *Staphylococcus*, the two most abundant genera in sample Even3T, show different length distributions. The dashed vertical lines in **(A)** indicate the average read lengths. The arrows in **(B)** indicate the fraction of reads from each taxon removed after discarding sequences shorter than $l_{min} = 250$ nt. **(C, D, E)**: Errors per base distributions of *Streptococcus* and *Staphylococcus* reads in the **(C)** raw dataset, **(D)** after trimming the reads to 250 nt and discarding the ones shorter than the cut-off, and **(E)** after collapsing the trimmed reads. The dashed lines indicate average errors per base. Note that length trimming substantially increases the difference between the *Streptococcus* and *Staphylococcus* error distributions **(D)** when compared to that of the raw reads **(C)**. Filtering at this point would cause a 56.2% overrepresentation of *Streptococcus* versus *Staphylococcus* (see text and Supplementary Note 2). Collapsing identical reads prior to filtering solves this problem **(E)**, reducing the overrepresentation to 1%. **(F)**: Compositional bias generated during the pre-processing and filtering of the six 454 mock community samples, measured as the Bray-Curtis dissimilarity between the raw and the processed read communities. This shows that results in **(C, D, E)** can be generalized to all the taxa present in all the samples.

any present or future sequencing platform, with no need for further modifications.

In practice, quality-score calling ultimately depends on the sequencing platform manufacturer (454 quality-scores, e.g., do not predict per-base error probabilities, but the probability of overestimating the homopolymer length), and its accuracy is also influenced by the choice of primers and library preparation methods (35,36). Nonetheless, we have shown that, for the three sequencing platforms studied in this work, Poisson binomial filtering was able to correctly discriminate between good and erroneous sequences based solely on quality score information.

The fact that our method only relies on quality scores means that it will only account for sequencing errors, but not other errors such as PCR substitutions. However, it has been described that sequencing errors are responsible for the majority of singletons generated in molecular ecology studies (15,17). PCR chimeras are other source of spurious diversity, but dedicated algorithms such as UCHIME are able to accurately detect a large majority of them (see 15 for a more thorough discussion on this subject).

During the course of this research, we have also focused on a source of taxonomic bias that may have affected the results of many molecular ecology studies. Notwithstanding the fact that upstream processes might introduce large biases in the amplicon pool to be sequenced, most of the methods used for filtering and analyzing marker-gene reads operate under the implicit (or even explicit, see 17) assumption that the probability of having j errors is the same for all reads, regardless of their origin. However, sequences from different taxa may have different length (for 454 and IonTorrent) and quality (for 454, IonTorrent and Illumina) distributions. This leads to the artificial enrichment of some taxa versus others during the quality filtering step, potentially compromising the quantitative interpretation of molecular ecology results obtained by high-throughput sequencing of marker-gene sequences. These biases are likely originated during base/quality calling: for instance, 454 reads show a systematic decrease in quality after homopolymer regions (37), which will penalize the taxa with longer homopolymer stretches on its 16S gene. We have nonetheless demonstrated that collapsing identical reads before the quality-filtering step greatly mitigates this issue.

In summary, the methodologies presented in this work substantially improve the existing filtering approaches in terms of OTU accuracy, observed alpha diversity and ob-

served community composition, delivering a more faithful representation of the original microbial communities present in the studied samples. Our algorithm is fast, easy to implement and works for every sequencing platform, constituting a valuable addition to all the existing pipelines for analyzing microbial ecology data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank G. Ackermann, N.A. Bokulich, J.G. Caporaso and J.R. Rideout for facilitating access to the MiSeq mock community datasets, S. Salipante for providing information on the IonTorrent mock community datasets, P.S. Schloss for his insights into mothur's implementation of the Needleman-Wunsch algorithm, and A. Arce-Rodríguez, M.J. Gómez, J. Iranzo, S. Lincoln, F. López de Saro, S. Manrubia and C. Pedrós-Alió for useful feedback and discussions of the draft manuscript.
Author Contributions: F.P.S. and J.A. conceived the algorithm. J.A. developed the analytical work. F.P.S. and V.P. conceived the experiments. F.P.S. implemented the algorithm and designed and performed the experiments. F.P.S., J.A. and V.P. participated in the motivation and discussion of the results and contributed to the writing of the manuscript.

FUNDING

Subdirección General de Proyectos de Investigación of Spanish Ministerio de Economía y Competitividad [AYA2011-24803, FIS2011-27569, FIS2014-57686]; Spanish Comunidad de Madrid [MODELICO-CM-S2009ESP-1691]; European Research Council [250350]; computational resources were provided by the Data Intensive Academic Grid, which is supported by the United States National Science Foundation [0959894]; Spanish Consejo Superior de Investigaciones Científicas (CSIC) [to F.P.S. JAE-pre fellowship]. Funding for open access charge: Subdirección General de Proyectos de Investigación of Spanish Ministerio de Economía y Competitividad [AYA2011-24803].
Conflict of interest statement. None declared.

REFERENCES

1. Roesch,L.F., Fulthorpe,R.R., Riva,A., Casella,G., Hadwin,A.K., Kent,A.D., Daroub,S.H., Camargo,F.A.O., Farmerie,W.G. and Triplett,E.W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.*, **1**, 283–290.
2. Costello,E., Lauber,C. L, Hamady,M., Fierer,N., Gordon,J.I. and Knight,R. (2009) Bacterial community variation in human body habitats across space and time. *Science*, **326**, 1694–1697.
3. Caporaso,J.G., Lauber,C.L., Walters,W.A., Berg-Lyons,D., Lozupone,C.A., Turnbaugh,P.J., Fierer,N. and Knight,R. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA*, **108**, 4516–4522.
4. Yatsusenko,T., Rey,F.E., Manary,M.J., Trehan,I., Dominguez-Bello,M.G., Contreras,M., Magris,M., Hidalgo,G., Baldassano,R.N., Anokhin,A.P. *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–227.
5. Ding,T. and Schloss,P. (2014) Dynamics and associations of microbial community types across the human body. *Nature*, **509**, 357–360.
6. Hugoni,M., Taib,N., Debroas,D., Domaizon,I., Jouan Dufournel,I., Bronner,G., Salter,I., Agogué,H., Mary,I. and Galand,P.E. (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc. Natl. Acad. Sci. USA*, **110**, 6004–6009.
7. Kunin,V., Engelbrektson,A., Ochman,H. and Hugenholtz,P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.*, **12**, 118–123.
8. Schloss,P., Westcott,S.L., Ryabin,T., Hall,J.R., Hartmann,M., Hollister,E.B., Lesnievski,R.A., Oakley,B.B., Parks,D.H., Robinson,C.J. *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, **75**, 7537–7521.
9. Caporaso,J.G., Kuczynski,J., Stombaugh,J., Bittinger,K., Bushman,F., Costello,E.K., Fierer,N., González-Peña,A., Goodrich,J.K., Gordon,J.I. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336.
10. Edgar,R. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
11. Huang,Y., Niu,B.F., Gao,Y., Fu,L.M. and Li,W.Z. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.
12. Quince,C., Lanzen,A., Davenport,R.J. and Turnbaugh,P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.
13. Schloss,P., Gevers,D. and Wescott,S. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLOS One.*, **6**, e27310.
14. Caporaso,J.G., Lauber,C.L., Walters,W.A., Berg-Lyons,D., Huntley,J., Fierer,N., Owens,S.M., Betley,J., Fraser,L., Bauer,M. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.*, **6**, 1621–1624.
15. Edgar,R. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*, **10**, 996–998.
16. Preheim,S.P., Perrotta,A.R., Martin-Platero,A.M., Gupta,A. and Alm,E.J. (2013) Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol*, **79**, 6593–6603.
17. Tikhonov,M., Leach,R.W. and Wingreen,N.S. (2014) Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J.*, epub ahead of print 11 July 2014, doi:10.1038/ismej.2014.117.
18. Quince,C., Lanzén,A., Curtis,T.P., Davenport,R.J., Hall,N., Head,I.M., Read,L.F. and Sloan,W.T. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods*, **6**, 639–641.
19. Reeder,J. and Knight,R. (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat. Methods*, **7**, 668–669.
20. Kozich,J., Westcott,S., Baxter,N., Highlander,S. and Schloss,P. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.*, **79**, 5112–5120.
21. Bokulich,N., Subramanian,S., Faith,J.J., Gevers,D., Gordon,J.I., Knight,R., Mills,D.A. and Caporaso,J.G. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods*, **10**, 57–59.
22. Butler,K. and Stephens,M. (1993) The distribution of a sum of binomial random variables. Tech. Rep. Stanford University 467, Stanford, USA.
23. Hildebrand,F., Tadeo,R., Voigt,A.Y., Bork,P. and Raes,J. (2014) LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome*, **2**, 30.
24. Haas,B.J., Gevers,D., Earl,A.M., Feldgarden,M., Ward,D.V. and Giannoukos,G. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.*, **21**, 494–504.
25. Quast,C., Pruesse,E., Yilmaz,P., Gerken,J., Schweer,T., Yarza,P., Peplies,J. and Glöckner,F.O. (2013) The SILVA ribosomal RNA gene

database project: improved data processing and web-based tools. *Nucleic Acids Res.*, **41**, 590–596.
26. Edgar,R.C. and Flyvbjerg,H. (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, **31**, 3476-3482, btv401.
27. Edgar,R., Haas,B., Clemente,J., Quince,C. and Knight,R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.
28. Gaspar,J. and Thomas,W. (2013) Assessing the consequences of denoising marker-based metagenomic data. *PLOS One*, **8**, e60458.
29. Klappenbach,J.A., Dunbar,J.M. and Schmidt,T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.*, **66**, 1328–1333.
30. Polz,M. and Cavanaugh,C. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.*, **64**, 3724–3730.
31. Klindworth,A., Pruesse,E., Schweer,T., Peplies,J., Quast,C., Horn,M. and Glöckner,F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.*, **41**, e1.
32. Ross,M., Russ,C., Costello,M., Hollinger,A., Lennon,N.J., Hegarty,R., Nusbaum,C. and Jaffe,D.B. (2013) Characterizing and measuring bias in sequence data. *Genome biol.*, **14**, R51.
33. Turnbaugh,P.J., Quince,C., Faith,J.J., McHardy,A.C., Yatsunenko,T., Niazi,F., Affourtit,J., Egholm,M., Henrissat,B., Knight,R. *et al.* (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl. Acad. Sci. USA*, **107**, 7503–7508.
34. Salipante,S.J., Kawashima,T., Rosenthal,C., Hoogestraat,D.R., Cummings,L.A., Sengupta,D.J., Harkins,T.T., Cookson,B.T. and Hoffman,N.G. (2014) Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.*, **80**, 7583–7591.
35. Loman,N.J., Misra,R.V., Dallman,T.J., Constantinidou,C., Gharbia,S.E., Wain,J. and Pallen,M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.*, **30**, 434–439.
36. Schirmer,M., Ijaz,U.Z., D'Amore,R., Hall,N., Sloan,W.T. and Quince,C. (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic acids res.*, **43**, e37.
37. Brockman,W., Alvarez,P., Young,S., Garber,M., Giannoukos,G., Lee,W.L., Russ,C., Lander,E.S., Nusbaum,C. and Jaffe,D.B. (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res.*, **18**, 763–770.